②

Japanese Patent Laid-open Publication No.: 2003-163687 A

Publication date : June 6, 2003

Applicant : NIPPON TELEGRAPH & TELEPHONE CORP

Title : METHOD AND DEVICE FOR CONTROLLING ROUTE

(54) [Title of the Invention] METHOD AND DEVICE FOR CONTROLLING ROUTE

(57) [Abstract]

[Object] To ensure dynamic optimization of distribution of traffic in a plurality of paths in a network.

[Solution] A flow classifying unit 1 classifies flows of input packets into respective groups based on preset weighting factors thereof, and flow output units 2, 3 output the flows to different paths for the groups. A bandwidth measuring unit 4 measures that of bandwidths used in the bandwidths which is used by each flow group by group, and a unit bandwidth calculator 5 calculates a unit use bandwidth for each group by dividing the measured use bandwidth of each group by a weighting factor previously assigned to the group. A weighting factor adjusting unit 6 adjusts the weighting factor of each group based on the unit use bandwidth of each group.

[Scope of Claims for Patent]

[Claim 1] A path control method for use in a path control apparatus that is connected to a packet transfer network, identifies a flow to which an input packet belongs from an attribute thereof, and outputs the packet to one of paths with the flow as a unit, comprising:

a step of classifying a flow of an input packet to one of groups based on a preset weighting factor of each group;

a step of outputting flows of the groups to different paths for the groups;

a step of measuring bandwidths used in the paths group by group; and

a step of adjusting the weighting factor of each group based on a ratio of the measured bandwidths.

[Claim 2] The path control method according to claim 1, further comprising a step of calculating a unit use bandwidth for each of the groups by dividing each measured bandwidth by a weighting factor previously assigned to the group, and

wherein at a time of adjusting the weighting factors, the weighting factor of each group is adjusted based on each unit use bandwidth.

[Claim 3] The path control method according to claim 2, wherein at a time of adjusting the weighting factors, the weighting factors are changed in a way as to average the unit use bandwidth of each group.

[Claim 4] The path control method according to claim 2, wherein at a time of adjusting the weighting factors, when the unit use bandwidth of the group is greater than the bandwidth of another group, the weighting factor of the former group is increased, and when the unit use bandwidth of the group is smaller than the bandwidth of another group, the weighting factor of the former group is decreased.

[Claim 5] The path control method according to claim 1, wherein at a time of adjusting the weighting factors, the weighting factors are adjusted in a way that a ratio of the weighting factors of the groups becomes equal to a ratio of the bandwidths.

[Claim 6] The path control method according to claim 1, wherein at a time of classifying the flows, classification is performed in a way that the number of flows to be classified into each group is proportional to the associated weighting factor.

[Claim 7] The path control method according to claim 1, wherein at a time of classifying the flows, a connection of a fourth layer of an OSI reference model is used as each flow.

[Claim 8] The path control method according to claim 1, wherein at a time of classifying the flows, a sequence of input packets having at least one of a sender address or a destination address of a third layer of an OSI reference model identical are taken as a same flow.

[Claim 9] The path control method according to claim 1, wherein at a time of classifying the flows, hash values of input packets are calculated from a hash function including at least one of a sender address, a destination address, a sender port number and a destination port number of an input packet as an argument, and those input packets whose hash values obtained are identical are classified into a same group.

[Claim 10] The path control method according to claim 1, wherein at a time of classifying the flows, a flow storage table which stores flows of classified input packets is used, when a new input packet belongs to any flow stored in the flow storage table, the input packet is classified into the same group as the flow, and when the input packet belongs to a new flow which is not stored in the flow storage table, the new flow is classified into that of the groups whose weighting factor is to be increased.

[Claim 11] The path control method according to claim 1, wherein as the path, a path which is distinguished based on LSP of MPLS is used.

[Claim 12] The path control method according to claim 1, wherein as the path, a path which is distinguished based on a virtual connection of PPP is used.

[Claim 13] The path control method according to claim 1, wherein as the path, a path which is distinguished based on a virtual connection VC or a virtual path VP of ATM is used.

[Claim 14] The path control method according to claim 1, wherein as the path, a path which is distinguished based on VLAN defined in IEEE 802.1Q is used.

[Claim 15] The path control method according to claim 1, wherein as the path, a path which is distinguished based on a node of a next hop of the path control apparatus is used.

[Claim 16] A path control apparatus that is connected to a packet transfer network, identifies a flow to which an input packet belongs from an attribute thereof, and outputs the packet to one of paths with the flow as a unit, comprising:

a flow classifying unit that classifies a flow of an input packet to one of groups based on a preset weighting factor of each group;

a flow output unit that outputs flows of the groups to different paths for the groups;

a bandwidth measuring unit that measures bandwidths used in the paths group by group; and

a weighting factor adjusting unit that adjusts the weighting factor of each group based on a ratio of the measured bandwidths of the groups measured by the bandwidth measuring unit.

[Claim 17] The path control apparatus according to claim 16, further comprising a unit bandwidth calculator that

calculates a unit use bandwidth for each of the groups by dividing each bandwidth measured by the bandwidth measuring unit by a weighting factor previously assigned to the group, and

wherein the weighting factor adjusting unit adjusts the weighting factor of each group based on each unit use bandwidth calculated by the unit bandwidth calculator.

[Claim 18] The path control apparatus according to claim 17, wherein the weighting factor adjusting unit changes the weighting factors in a way as to average the unit use bandwidth of each group.

[Claim 19] The path control apparatus according to claim 17, wherein when the unit use bandwidth of the group is greater than the bandwidth of another group, the weighting factor adjusting unit increases the weighting factor of the former group, and when the unit use bandwidth of the group is smaller than the bandwidth of another group, the weighting factor adjusting unit decreases the weighting factor of the former group.

[Claim 20] The path control apparatus according to claim 16, wherein the weighting factor adjusting unit adjusts the weighting factors in a way that a ratio of the weighting factors of the groups becomes equal to a ratio of the bandwidths.

[Claim 21] The path control apparatus according to claim 16, wherein the flow classifying unit performs classification in a way that the number of flows to be classified into each group is proportional to the associated weighting factor.

[Claim 22] The path control apparatus according to claim 16, wherein the flow classifying unit uses a connection of a fourth layer of an OSI reference model as each flow.

[Claim 23] The path control apparatus according to claim 16, wherein the flow classifying unit takes a sequence of input packets having at least one of a sender address or a destination address of a third layer of an OSI reference model identical as a same flow.

[Claim 24] The path control apparatus according to claim 16, wherein the flow classifying unit includes a hash calculator that calculates hash values of input packets from a hash function including at least one of a sender address, a destination address, a sender port number and a destination port number of an input packet as an argument, and a hash value group classifying unit that classifies those input packets whose hash values obtained are identical into a same group.

[Claim 25] The path control apparatus according to claim 16, wherein the flow classifying unit has a flow storage table which stores flows of classified input packets, when a new input packet belongs to any flow stored in the flow storage table, the flow classifying unit classifies the new input packet into the same group as the flow, and when the input packet belongs to a new flow which is not stored in the flow storage table, the flow classifying unit classifies the new flow into that of the groups whose weighting factor is to be increased.

[Claim 26] The path control apparatus according to claim 16, wherein as the path, a path which is distinguished based on LSP of MPLS is used.

[Claim 27] The path control apparatus according to claim 16, wherein as the path, a path which is distinguished based on a virtual connection of PPP is used.

[Claim 28] The path control apparatus according to claim 16, wherein as the path, a path which is distinguished based on a virtual connection VC or a virtual path VP of ATM is used.

[Claim 29] The path control apparatus according to claim 16, wherein as the path, a path which is distinguished based on VLAN defined in IEEE 802.1Q is used.

[Claim 30] The path control apparatus according to claim 16, wherein as the path, a path which is distinguished based on a node of a next hop of the path control apparatus is used.

[Detailed Description of the Invention]

[0001]

[Technical Field of the Invention]

The present invention relates to a path control method and apparatus, and more particularly, to a path control method and apparatus that dynamically distribute a load to a plurality of paths in a packet transport network.

[0002]

[Conventional Technique]

Although there is a significant improvement on the speed of an access line in recent communication environment, i.e., the transmission speed between a service subscriber terminal and a nearby subscriber holding apparatus, an improvement on the speed of the core network of a carrier is delayed due to various factors, and the bandwidths of a network core can easily be used up by the traffic of a small number of users.

[0003] To improve the utilization factor of network resources in such a situation, the field of traffic engineering (hereinafter, TE: Traffic Engineering) has been studied actively in recent years. While the packet transfer along the path (the shortest path) which involves the lowest cost at an end point was the main stream in the conventional IP network, the TE aims at avoiding that a load focuses on a specific relay line, thereby improving the utilization factor of network resources by allowing the

path with the second and third lowest costs to participate in packet transfer.

[0004] In the TE, for example, the necessary bandwidth of the circuit that connects each base of a user is estimated by the measurement on the network side, or the demand from the user, and calculation for optimization of the path to be assigned to the circuit is performed so that network resources can be used most efficiently. However, this calculation is complicated, and takes a long time to acquire the optimal solution in a large network. As a result, it is not rare that changing the setting of a path accompanying a change in traffic takes several days or more. The method of optimizing the network resources over a long period of time is hereinafter called as static TE.

[0005] The traffic with a high burst to multiple unspecified persons is rapidly increasing as traffic of the Internet due to the aforementioned improvement on the speed of the access line. That is, unlike the conventional dedicated line or a telephone, a communication party is not fixed and what is more, there is a large difference in use bandwidth between the time of data transfer and the time of not transferring data. With the conventional static TE alone, therefore, it is difficult to efficiently use network resources for the traffic for which a path and a use bandwidth are changed sharply for a short time. There is a strong demand for a method which is sensitively responsive to a short-time change and optimizes resources. Such a method is called hereinafter as dynamic TE.

[0006]

[Problems to be Solved by the Invention]

However, such dynamic TE has a problem that the conventional load distribution method is not applicable. Next, the technical problem at the time of applying the

conventional load distribution method to the dynamic TE is explained. Execution of the TE generally involves three stages, (1) measuring the load of each path in a network, i.e., the degree of utilization, (2) analyzing the measuring result and performing calculation to optimize allocation of resources, and (3) transferring traffic through a new path based on the calculation result.

[0007] Measurement of the load (1) in an IP network is generally carried out by remotely acquiring information from a server that monitors load information with respect to each path control apparatus to be monitored using SNMP (Simple Network Management Protocol). However, applying this method to the dynamic TE has many technical problems. The following will describe problems at the time of collecting load information in case of using SNMP as an example. First, it is difficult to handle an overloaded link and a shaper. When an overloaded link (link which has used up the physical speed) or a shaper (apparatus that shapes traffic to a fixed bandwidth) is present in a path, information is insufficient if the flow rate of the link is merely acquired as load information. That is, with the maximum speed used up, information as to how much it is actually overloaded and how many traffics should be distributed to other paths is not acquired.

[0008] To solve this problem strictly, it is necessary to measure the difference between the input bandwidth and the output bandwidth of each path control apparatus with respect to each path. What is more, the path control apparatus generally has a plurality of input interfaces and output interfaces, and traffic repeats meeting and parting in a plurality of path control apparatuses, so that it is necessary to investigate not only the amount of traffic which merges after passing through all paths but also the

path control tables in the path control apparatuses in order to investigate the cause of the congestion at a certain link. It is however extremely difficult to collect and analyze those pieces of information at a high speed. Even an overloaded link or a shaper is present in a path, therefore, a simple and effective method of estimating the degree of load and the necessary amount of distribution to other paths is required.

[0009] Secondly, due to the provision of the server, a new complication is increased. Dynamic load distribution requires that load information be collected from time to time in a cycle of several seconds. In collecting load information frequently in a large network, it is more efficient for an exclusive server that gathers load information to collectively gather information than path control apparatuses individually performing measurement, analysis and the like. This however brings up a new complication to the configuration of a network. For example, it is necessary to take into consideration the setting of a separate circuit for load information, double circuit configuration therefor, double server configuration and the like. There is a strong demand for an efficient information collecting method which does not use such a server.

[0010] Thirdly, there is an increase in the number of management targets. To bypass each path appropriately, it is necessary to measure the load for every path with respect to all the path control apparatuses in a network. Since the number of paths and the number of path control apparatuses increase drastically in a large network, however, it is difficult to perform measurements at individual en-route points in these paths, and to analyze the results. This results in long calculation required for

optimization in response to a change in load, making adaptation to the dynamic TE difficult. A method of effectively performing load distribution with fewer management targets is required.

[0011] Fourthly, there is a difficulty in attending to a failure. The dynamic TE is demanded of the capability which spontaneously copes with a failure in a network. Therefore, when there is little traffic in a certain path, it is necessary to find quickly out whether the cause is a failure in a link or a failure in path control, or is simply from vacancy. When it is vacant, it is preferable to actively let flows to flow there, whereas when a failure has occurred in a path, it is desirable not to use the path.

[0012] However, it is difficult to determine such a failure using SNMP alone in an actual network. When the repeater and switch that do not have the management function of the SNMP are present on a path, for example, the failure produced in that part cannot be detected. Moreover, the cause of a failure is not just a failure in an interface, but various elements, such as generation of a loop in the transitional condition of path control, are possible. Therefore, accurate detection of a failure on path control requires acquisition of path control information held in each path control apparatus (information which is exchanged by the path control protocol) and a path control table in addition to monitoring the statuses of all the elements associated with packet transfer including such a repeater and a switch.

[0013] What is making the problems more difficult is that a plurality of path control protocols are used in the same network. That is, since it is common to use manual static setting of a path and setting of a path using dynamic path control protocols, such as OSPF (Open Shortest

Path First) and RIP (Routing Information Protocol), together, it is necessary to collect information for all of the static or dynamic path settings with the optimal method, and to perform mutual conversion thereof and systematic analysis thereof in order to detect a failure in path control. For example, it is necessary to acquire information representing a link status for OSPF, and check up/down information for adjoining path control apparatuses and from which adjoining apparatus a path is learned for RIP. Even when a failure occurs in a network, therefore, a method of detecting the failure promptly and optimizing a path is required.

[0014] As apparent from the above, there are many matters in applying the conventional load distribution method to the dynamic TE, and it is difficult to dynamically optimize paths in a network on the practical level. The present invention is to solve such a problem, and aims to provide a path control method and apparatus that can dynamically optimize distribution of traffic over a plurality of paths in a network.

[0015]

[Means for Solving Problem]

To achieve the object, a path control method according to the present invention is a path control method for use in a path control apparatus that is connected to a packet transfer network, identifies a flow to which an input packet belongs from an attribute thereof, and outputs the flow to one of paths with the flow as a unit, comprising: a step of classifying a flow of an input packet to one of groups based on a preset weighting factor of each group; a step of outputting flows of the groups to different paths for the groups; a step of measuring bandwidths used in the paths group by group; and a step of adjusting the weighting

factor of each group based on a ratio of the measured bandwidths.

[0016] A step of calculating a unit use bandwidth for each of the groups by dividing each measured bandwidth by a weighting factor previously assigned to the group can further be provided, and at a time of adjusting the weighting factors, the weighting factor of each group can be adjusted based on each unit use bandwidth.

[0017] At a time of adjusting the weighting factors, the weighting factors can be changed in a way as to average the unit use bandwidth of each group. In addition, when the unit use bandwidth of the group is greater than the bandwidth of another group, the weighting factor of the former group can be increased, and when the unit use bandwidth of the group is smaller than the bandwidth of another group, the weighting factor of the former group can be decreased. At a time of adjusting the weighting factors, the weighting factors can be adjusted in a way that a ratio of the weighting factors of the groups becomes equal to a ratio of the bandwidths.

[0018] At a time of classifying the flows, classification can be performed in a way that the number of flows to be classified into each group is proportional to the associated weighting factor. In addition, a connection of a fourth layer of an OSI reference model can be used as each flow, or a sequence of input packets having at least one of a sender address or a destination address of a third layer of an OSI reference model identical can be taken as a same flow.

[0019] At a time of classifying the flows, hash values of input packets can be calculated from a hash function including at least one of a sender address, a destination address, a sender port number and a destination port number

of an input packet as an argument, and those input packets whose hash values obtained are identical can be classified into a same group, or a flow storage table which stores flows of classified input packets can be used, when a new input packet belongs to any flow stored in the flow storage table, the input packet can be classified into the same group as the flow, and when the input packet belongs to a new flow which is not stored in the flow storage table, the new flow can be classified into that of the groups whose weighting factor is to be increased.

[0020] As the path, a path which is distinguished based on LSP of MPLS can be used. In addition, a path which is distinguished based on a virtual connection of PPP, a path which is distinguished based on a virtual connection VC or a virtual path VP of ATM, a path which is distinguished based on VLAN defined in IEEE 802.1Q, or a path which is distinguished based on a node of a next hop of the path control apparatus can be used.

[0021] A path control apparatus according to the present invention is a path control apparatus that is connected to a packet transfer network, identifies a flow to which an input packet belongs from an attribute thereof, and outputs the flow to one of paths with the flow as a unit, comprising: a flow classifying unit that classifies a flow of an input packet to one of groups based on a preset weighting factor of each group; a flow output unit that outputs flows of the groups to different paths for the groups; a bandwidth measuring unit that measures bandwidths used in the paths group by group; and a weighting factor adjusting unit that adjusts the weighting factor of each group based on a ratio of the measured bandwidths of the groups measured by the bandwidth measuring unit.

[0022] A unit bandwidth calculator that calculates a unit use bandwidth for each of the groups by dividing each bandwidth measured by the bandwidth measuring unit by a weighting factor previously assigned to the group can further be provided, and the weighting factor adjusting unit can adjust the weighting factor of each group based on each unit use bandwidth calculated by the unit bandwidth calculator.

[0023] The weighting factor adjusting unit can change the weighting factors in a way as to average the unit use bandwidth of each group. In addition, when the unit use bandwidth of the group is greater than the bandwidth of another group, the weighting factor adjusting unit can increase the weighting factor of the former group, and when the unit use bandwidth of the group is smaller than the bandwidth of another group, the weighting factor adjusting unit can decrease the weighting factor of the former group. The weighting factor adjusting unit can adjust the weighting factors in a way that a ratio of the weighting factors of the groups becomes equal to a ratio of the bandwidths.

[0024] The flow classifying unit can perform classification in a way that the number of flows to be classified into each group is proportional to the associated weighting factor. In addition, the flow classifying unit can use a connection of a fourth layer of an OSI reference model as each flow, or the flow classifying unit can take a sequence of input packets having at least one of a sender address or a destination address of a third layer of an OSI reference model identical as a same flow.

[0025] At a time the flow classifying unit classifies the flows, a hash calculator can calculate hash values of

input packets from a hash function including at least one of a sender address, a destination address, a sender port number and a destination port number of an input packet as an argument, and a hash value group classifying unit can classify those input packets whose hash values obtained are identical into a same group.

[0026] The flow classifying unit can have a flow storage table which stores flows of classified input packets, when a new input packet belongs to any flow stored in the flow storage table, the flow classifying unit can classify the new input packet into the same group as the flow, and when the input packet belongs to a new flow which is not stored in the flow storage table, the flow classifying unit can classify the new flow into that of the groups whose weighting factor is to be increased.

[0027] As the path, a path which is distinguished based on LSP of MPLS can be used. In addition, a path which is distinguished based on a virtual connection of PPP, a path which is distinguished based on a virtual connection VC or a virtual path VP of ATM, a path which is distinguished based on VLAN defined in IEEE 802.1Q, or a path which is distinguished based on a node of a next hop of the path control apparatus can be used.

[0028]

[Embodiments of the Invention]

Exemplarily embodiments of the present invention will be explained below in detail with reference to the accompanying drawings. Fig. 1 is a schematic configuration diagram of a packet transfer network to which a path control method according to one embodiment of the present invention is adapted. A packet transfer network N11 includes a single load balancer B11 having a path control apparatus that employs the path control method according to

one embodiment of the present invention, and three routers R11, R12, R13, which are connected to one another by transmitting units L11. The routers R11, R12, R13 are general units that transfer packets, and switches, hubs, repeaters, packet exchangers or the like can be used as the routers. While the use of three routers is taken as an example hereinafter, an arbitrary number of routers can be used.

[0029] A plurality of terminals T11, T12 are connected to the load balancer B11 via the transmitting units L11. While the use of one load balancer B11 is taken as an example hereinafter, an arbitrary number of load balancers can be used. Although the load balancer B11 is directly connected to the terminals T11, T12 by transmitting units L11, an arbitrary number of routers or networks can be present between both. The load balancer B11 does not need to be provided at an entry point of the network as in Fig. 1, but can be incorporated in, for example, the terminals T11, T12 or the routers R11, R12.

[0030] Not all of the transmitting units L11 should be identical, and different techniques can be used for the transmitting unit that connects T11 to B11 and for the transmitting unit that connects B11 to R11. In addition, although not shown, a very large number of terminals can be connected to the packet transfer network N11 via other load balancers and routers, and the amount of traffic that flows through the individual parts in the network N11 can vary from time to time.

[0031] Fig. 2 is a concept diagram of a plurality of paths in a case that packets are transferred to a terminal T13 via the load balancer B11 with the terminals T11 and T12 as senders. In Fig. 1, there are multiple paths possible that run toward the terminal T13 from the load

balancer B11 via the routers R11, R12, R13. For easier understanding, a case that two packets paths P11, P12 are used will be considered. A path of load balancer B11 → router R11 → router R13 → terminal T13 is used as the path P11, and a path of load balancer B11 → router R12 → router R13 → terminal T13 is used as the path P12.

[0032] The packets output from the paths P11, P12 terminals T11, T12 reach the terminal T13 passing through either the path P11 or the path P12 while repeating meeting and parting to and from traffic from other terminals. As the loads on the routers R11, R12 and on the transmitting units L11 in the path change from time to time, the target of the load balancer B11 is to transfer traffic to the paths P11, P12 in a way that the utilization factor of the network resources becomes highest according to the present invention the loads on the paths.

[0033] Fig. 3 is a block diagram of a configuration example of the load balancer B11. The load balancer B11 includes a flow classifying unit 1, flow output units 2, 3, a bandwidth measuring unit 4, a unit bandwidth calculator 5, and a weighting factor adjusting unit 6. The load balancer B11 manages packets having the same attribute as a single flow based on information which uses the attributes of a packet to be input, such as a sender address, a destination address, a sender port number, and a destination port number, singularly or in combination. The load balancer B11 controls distribution of loads on the paths by classifying those flows into a plurality of groups, and adjusting the bandwidths of the paths to be used in the groups.

[0034] The flow classifying unit 1 is a functional unit that classifies flows to which input packets belong into

individual groups based on weighting factors preset for respective groups. The flow output unit 2, 3 is a functional unit that is provided for the respective path P11, P12 and outputs flows, classified into each group by the flow classifying unit 1, to the associated path group by group. The bandwidth measuring unit 4 is a functional unit that measures the bandwidth the group uses in each path.

[0035] The unit bandwidth calculator 5 is a functional unit that divides the use bandwidth of each group measured by the bandwidth measuring unit 4 by the weighting factor of the group to calculate the unit use bandwidth per weighting factor of each group. The weighting factor adjusting unit 6 is a functional unit that adjusts the weighting factor of each group in a way that the unit use bandwidth of each group calculated by the unit bandwidth calculator 5 is averaged. Each of those functional units can be configured by cooperation of hardware that includes a microprocessor like a CPU, and a peripheral circuit, and software that is executed by the microprocessor, or can be configured only by hardware circuits.

[0036] As an example, the following will describe a case that flows are classified into two groups G11 and G12, flows of the group G11 are output to the path P11 from the flow output unit 2, and flows of the group G12 are output to the path P12 from the flow output unit 3.

[0037] Fig. 4 is a flowchart of a path control process in the load balancer. Packets input from the terminals T11, T12 are classified into the groups G11, G12 based on the weighting factors of the groups (step 100). The classification method is such that, for example, the number of flows to be classified into each group is proportional to the weighting factor. Fig. 3 is an example where the

number of input flow is 6, and weighting factors W11, W12 of the groups G11, G12 are 1 and 2, respectively. This classification method classifies two of the input flows into the group G11, and the other four into the group G12.

[0038] The groups G11, G12 are respectively output to the paths P11, P12 by the flow output units 2, 3 (step 101). The unit use bandwidths of the flows of the packets output to the paths P11, P12 are measured from time to time for each group by the bandwidth measuring unit 4 (step 102). Based on the measured results, the bandwidth per unit weighting factor is calculated (step 103). When the measured unit use bandwidths in the groups G11, G12 are respectively 10 and 12, for example, dividing the values by the corresponding weighting factors 1 and 2 yields the unit use bandwidths of 10 and 6 per unit weighting factor, respectively.

[0039] Based on the unit use bandwidth for each group calculated by the unit bandwidth calculator 5, the weighting factor adjusting unit 6 adjusts the weighting factor of each group (step 104). Therefore, when the flow classifying unit 1 reclassifies the flows or classifies new flows thereafter, the process returns to step 100 where the weighting factor adjusted by the weighting factor adjusting unit 6 will be used, thus averaging the bandwidths to be used by flows passing through each path. This dynamically optimizes the distribution of traffic in a plurality of paths in the network. Accordingly, the utilization factor of the network resources for traffic with high burst is improved, making it possible to provide many clients with a low-cost packet transfer capability.

[0040] The present invention has an aspect of providing a packet transfer method which dynamically changes the transfer path according to the present invention

corresponding to the load condition of the network. There is a specific example of a case that paths belonging to the same TCP connection are considered as flows. The traffics are classified in a way that the ratio of the numbers of TCP connections included in the groups becomes equal to the ratio of the numbers of weighting factors assigned to the groups, are respectively sent to different paths, the use bandwidths of the groups sent to the paths are measured group by group, and the measured use bandwidths are divided by the weighting factor of the group, thereby optimizing the number of weighting factors in a way that the unit use bandwidths per weighting factor among the groups become equal. Accordingly, even when a link or a shaper which uses up the physical bandwidth is present on a path, the throughputs of the TCP sessions that pass through the paths are quantitatively compared with each other to be optimized. This can overcome the first problem.

[0041] As each load balancer can execute the measurement and the calculation for optimization independently without using information about a load or a failure from other load balancers or routers, providing a server that is the second problem is unnecessary. As each load balancer can internally measure and process the use bandwidth of each group only for those flows which pass the load balancer, it is unnecessary to collect information on all pass points of all the paths over the entire network. The number of measurement targets becomes smaller as compared with the conventional method, and the measuring process is distributed to multiple load balancers, thereby overcoming the third problem.

[0042] The method also overcomes the fourth problem. This utilizes that a transport layer protocol, such as TCP, generally has a congestion control function. For example,

a terminal that performs communications using TCP dynamically changes the use bandwidth of a connection according to the quality of the network. When a failure or congestion occurs in the communication path, the transmission speed is automatically reduced, whereas when the communication path is empty, the transmission speed is automatically increased.

[0043] The present invention, unlike conventional examples, does not measure the use bandwidth of each path, but calculates the use bandwidth per unit weighting factor or the unit use bandwidth, which is a use bandwidth per TCP connection. When the unit use bandwidth among multiple paths are compared with one another and the unit use bandwidth is extremely small, it is considered that some sort of failure has occurred, and when the unit use bandwidth is increased, it is understood that the path is empty. This makes it possible to grasp whether a decrease in traffic in a path is originated from a failure in link or simply from the path being empty, so that distribution can be done in a short period of time.

[0044] Although the bandwidth measuring unit 4 measures the use bandwidth based on the outputs of the flow output units 2, 3, the use bandwidth can be measured from the output of the flow classifying unit 1 because the groups G11, G12 correspond in one to one to the paths P11, P12. Although the path control process has been explained as a sequence of flow processes referring to Fig. 4, the individual steps in the path control process can be executed singularly. Particularly, the measuring interval in the bandwidth measuring unit 4 can be periodic or can be changed intentionally. Alternatively, the measurement can be executed as needed.

[0045] The weighting factor adjusting method in the weighting factor adjusting unit 6 is explained with reference to Fig. 5. Fig. 5 is an explanatory diagram of an adjustment example for weighting factors. Given that the weighting factors of the groups G11, G12 are weighting factors W11 and W12, and the unit use bandwidths per unit weighting factor are BW11 and BW12, the weighting factor adjusting unit 6 changes the weighting factors W11, W12 according to the sizes of the unit use bandwidths BW11, BW12. When BW11>BW12, W11 is increased and W12 is decreased, and when BW11<BW12, W11 is decreased and W12 is increased. When BW11 is equal to BW12, the values of W11 and W12 does not need to be changed.

[0046] The adjustment of the weighting factors W11, W12 in a way as to be proportional to the number of flows included in each group G11, G12 increases the number of flows that pass through the path P11 and decreases the number of flows that pass through the path P12. Therefore, the system changes in the direction of making the bandwidth per flow uniform. With regard to the weighting factor adjusting method, the degrees of adjustment of the weighting factors W11, W12 can be changed according to the levels of the absolute values of unit use bandwidths BW11, BW12 or the level of the difference therebetween.

[0047] In the above description, a concept of a unit use bandwidth per unit weighting factor with respect to each path has been introduced. However, when control to make the unit use bandwidth of each path uniform is executed in step 104 (see Fig. 4) of adjusting the weighting factor, it is possible to take a structure that eliminates step 103 of calculating the unit use bandwidth per weighting factor. In this case, in step 104, taking the use bandwidth of each path measured in step 102 as an input variable, the

weighting factors should be adjusted in a way that the ratio of the weighting factors for the paths becomes equal to the ratio of the use bandwidth of the path.

[0048] The use of this structure can likewise eliminate the unit bandwidth calculator 5 in Fig. 3 in which case the weighting factor adjusting unit 6 receives the use bandwidths of the paths, measured by the bandwidth measuring unit 4, as inputs, and outputs those weighting factors which are equal to the ratio of the use bandwidths of the paths. By way of a specific example for explanation, when the measured use bandwidths of the paths P11, P12 are respectively 10 and 12, for example, connections should be classified by the ratio of 10:12 with respect to the paths P11, P12 to equalize the bandwidth per connection.

[0049] It is to be noted however that the movement of connections between paths this way can cause the use bandwidths in P11 and P12 to change from the initial 10:12, or can make the system unstable. Therefore, adjustment of weighting factors can be gradually changed in multiple separate times according to the use bandwidths that are acquired from time to time. Alternatively, for the purpose of accelerating the convergence speed, the amount of a change in weighting factor can be intentionally increased in consideration of a change in bandwidth after changing the weighting factor.

[0050] The flow classification method in the flow classifying unit 1 will be explained next with reference to Figs. 6 and 7. Figs. 6 and 7 are examples of the flow classification method. In the flow classification method in Fig. 6, the fourth layer connection of the OSI reference model is taken as a flow. As an attribute for classification, TCP (Transmission Control Protocol) is used as an example of the fourth layer protocol. In this case,

one connection of the TCP corresponds to one flow. The TCP has a congestion control function and a flow control function, and given that the amount of data and external conditions, such as delay and packet loss, are the same, the bandwidths of the individual connections tend to be equal. Therefore, adjusting W11, W12 in a way as to equalize BW11 and BW12 in the above-described manner changes the system in the direction of averaging all the bandwidths of the TCP connections that pass through both paths P11, P12. This improves the resource utilization efficiency over the entire network.

[0051] In the flow classification method in Fig. 7, a sequence of packets in which at least the sender or destination address of the third layer of the OSI reference model is the same is taken as a flow. As an attribute for classification, IP is used as an example of the third layer. In this case, the load for identifying flows can be made smaller as compared with the way of identifying flows according to the fourth layer protocol. This is because while four numerals, a sender IP address, a destination IP address, a sender port number and a destination port number, are generally needed to identify a TCP connection, the case of Fig. 7 requires the sender IP address only. In this example, the system changes in a way that the bandwidth per sender address becomes uniform in two paths. When TCP is used as the fourth layer protocol and sender IP addresses use about the same number of TCP connections on average, for example, the bandwidths to be used by terminals of the sender IP addresses become equal.

[0052] The methods shown in Figs. 6 and 7 need to check to which flow belongs every packet that passes through the load balancer based on the attribute of the packet, and store it. However, the essential requirement in the

embodiment is to classify input flows to G11 and G12 by the ratio proportional to the ratio of weighting factors, W11:W12. To achieve the object, a hash function can be used. Fig. 8 is an example of the configuration of the flow classifying unit 1 that uses a hash function which has the sender address, the destination address, the sender port number and the destination port number as arguments and integers of 1 to 9 as a value range, and includes a hash value calculator 21 and a hash value group classifying unit 22.

[0053] An available hash function is such that the numbers of flows included in individual minute portions of the value of the hash function become equal to one another. One example of such a hash function is a function of "(sender address + destination address + sender port number + destination port number) mod9 + 1" (AmodB indicating the remainder of A divided by B). When the hash value calculator 21 calculates the hash function for the packets, integers of 1 to 9 are obtained as hash values. In case of the Internet that has a sufficiently large number of flows, the numbers of flows that take the individual hash values become approximately equal. To finely adjust the proportion of flow classification, the number of divisions in the function, 9, should be made larger. The hash value group classifying unit 22 divides the value range to the previous ratio of W11:W12 to be respectively associated with the groups G11, G12. This can allow input flows to be classified by an arbitrary ratio as a consequence. The method has an advantage that it is unnecessary to store flows which pass through the load balancer.

[0054] Distinction of paths is explained below. The present embodiment requires that a plurality of paths be set for the same destination in the packet transfer network

N11. In the conventional Internet, because packets for the same destination pass through the same path, paths should be distinguished by somehow. As an example of a method available for the purpose, paths can be distinguished based on LSP (Label Switched Path) of MPLS (Multi Protocol Label Switching), a virtual connection of PPP (Point-to-Point Protocol), a virtual connection of a frame relay, a virtual path VP (Virtual Path) of ATM (Asynchronous Transfer Mode), a virtual channel VC (Virtual Connection), VLAN (Virtual Bridged Local Area Networks) of IEEE 802.1Q, a node (physical interface) of a next hop, a radio channel or frequency, the wavelength of WDM (Wavelength Division Multiplexing) or the like, however, the method is not restrictive as long as a plurality of paths can be distinguished.

[0055] Fig. 9 depicts a case that paths are distinguished by the LSP of an MPLS network. The MPLS network generally comprises a label edge router (Label Edge Router/LER) and a label switching router (Label Switching Router/LSR) connected to each other. The label edge router is provided at the outer circumferential portion of the MPSL network, and serves to receive packets from an external network or terminal and transfer the packets into the MPLS network, and transfer packets received from inside the MPLS network to an external network or terminal. The label switching router is located inside the MPLS network, and serves to transfer packets received from the label edge router or another label switching router to another label edge router or label switching router.

[0056] The LSP that is the path to which packets are transferred in the MPLS network is set up along with a series of label edge routers and label switching routers with the label edge router (Label Edge Router/LER) at the

entry point of the MPLS network inlet port as the starting point, and the label edge router at the exit of the MPLS network as the termination point. Generally, packet transfer acquires information, such as an input interface and a destination IP address, of a packet, searches a path control table based thereon in a label edge router, and classifies the packet into each transfer class (Forwarding Equivalence Class/FEC).

[0057] Next, after searching for the interface that should output the value of the label corresponding to each transfer class, and a packet, the second layer address of the OSI reference model of the label switching router of the next hop or the like, and encapsulating the packet by the MPLS header and the second layer header based on those information, the packet is transferred to the corresponding label switching router of the next hop. What is necessary is to perform load distribution in the MPLS network using the present invention, to install the load balancer that has the path control apparatus according to the present invention at, for example, the entry point of the MPLS network, and to set a plurality of label switching paths which have the apparatus as the starting point with respect to a transfer class which is to be subject to load distribution. Accordingly, the packet classified into an arbitrary transfer class can be classified and transferred to the group corresponding to each label switching path based on the weighting factor thereof.

[0058] Although the example which associates a plurality of label switching paths with one transfer class has been explained above, it is possible to associate one label switching path with one transfer class. In this case, each group and a transfer class are associated with each other in 1 to 1, and the group classification can also be

performed, at the same time, based on a weighting factor in the step of classifying an input packet into a transfer class.

[0059] The location at which the load balancer is installed is not restricted to the entry point of the MPLS network. That is, it is possible to perform load distribution among a plurality of label switching paths extending to the outlet label edge router with the load balancer as the starting point by placing the load balancer in that label switching path which starts at an arbitrary entry label edge router and ends at the outlet label edge router, setting a plurality of label switching paths which have the load balancer in the middle of the label switching path which extends to the outlet label edge router with the load balancer as the starting point. Accordingly, when the quality of an arbitrary label switching path deteriorates in the middle of a path, it becomes possible to assign a flow to another label switching path automatically, and the network utilization factor can always be kept high.

[0060] Fig. 10 depicts a case that a path is distinguished through the virtual connection of PPP. PPP is the protocol of the second layer of an OSI reference model, and, the virtual connection generally connects two adjoining nodes. When an arbitrary node is the end of a plurality of PPP connections, it is generally possible to clearly identify the packet that belongs to each virtual connection. Namely, when transmission media differ for every virtual connection, or when a plurality of virtual connections are set up through the same transmission medium like PPPoE (PPP over Ethernet (registered trademark)), a unique set of (a sender MAC Address, a destination MAC Address, and a session identifier) can be assigned for every virtual connection.

[0061] The case that one customer connects to a plurality of service providers simultaneously, and connects to a network therethrough as an example of the load distribution which uses such a PPP virtual connection will be explained as an example. To simplify the explanation, it is given on the assumption that a customer and each service provider are connected through one PPP virtual connection, respectively. That is, the load balancer that a customer owns has been connected to the remote access servers of a plurality of service providers through different PPP virtual connections, respectively.

[0062] In the above case, one group can be assigned to each service provider, and the corresponding ratio of weighting factors can be adjusted according to the use bandwidths toward the respect providers. That is, when one service provider is crowded and the service provider of another side is vacant, the flow which goes to the crowded service provider can be automatically assigned to the vacant service provider. Accordingly, it is possible to efficiently transfer packets to service providers even under the situation where the congestion of service providers changes from time to time.

[0063] Fig. 11 depicts a case that the VP or VC of the ATM is used. The VP or VC of the ATM can be used to connect, for example, between routers. Following is an explanation, as an example, of the configuration where two paths of going to an arbitrary but same destination from the load balancer exist, and are connected with the routers of the next hops of the paths by different VPs or VCs, respectively. In the case, the traffic which goes to the same destination in the load balancer can be classified into two groups, and the weighting factors for the

respective groups can be adjusted based on the result of measuring the respective use bandwidths.

[0064] Even if a failure occurs in either the VP or the VC, and arrival at the same destination becomes impossible, the weighting factor for the other one of the VP and VC increases automatically, so that a flow can be moved from one path which has the failure to the other path which has no failure. The same destination is not necessarily restricted to the last arrival point of a packet, but the point where a plurality of paths to the network which are to be passed en route are present can be considered as the same destination.

[0065] Fig. 12 depicts a case that a path is distinguished using the VLAN of IEEE 802.1Q. For example, when a packet can reach the same destination via a plurality of VLANs, each VLAN can be taken as one path. In this case, it becomes possible to transfer traffic, efficiently using each VLAN by assigning a weighting factor to each path, measuring the traffic which passes through each VLAN, and adjusting each weighting factor based on the result.

[0066] Fig. 13 depicts a case that a path is distinguished based on the node of the next hop (when interfaces differ physically). A node herein points out a general thing with the function of transmitting, receiving or transferring a packet based on the information on a network layer (or data link layer). The next hop is a node to be passed next as observed from an arbitrary node on the path which transfers a packet to the destination from the sender. As an example, the following will describe a case that the router of the service provider is connected to a plurality of other service providers through IX (Internet eXchange).

[0067] When a plurality of service provides are selectable as the next hop for a packet to arrive at the same destination network, it is possible to perform load distribution on the path that goes through the service providers using the path control method according to the present invention. When the structure of the IX is such that the routers of the individual providers are connected around the core switch of a data link layer, the IP address of the node of the next hop and the data link layer address corresponding thereto can be used as specific means for identifying a path.

[0068] By classifying packets addressed to the destination network into groups of a plurality of paths, measuring the use bandwidths for the groups, and reflecting the measurements on the values of weighting factors, when the quality of each path changes with time, it is possible to transfer traffic always using the resource of each path effectively. For example, when the quality of the path which goes through an arbitrary service provider degrades, the flow passing therethrough can be promptly assigned to the path which goes through other service providers automatically. On the contrary, when the path which goes through an arbitrary service provider is less busy than other paths, a flow can be automatically assigned to the less-busy path from the other paths.

[0069] When a failure occurs in the provider in the middle of an arbitrary path and arrival at a destination network is stopped, a congestion-control function works at the connection of the TCP which goes through the path and the throughput deteriorates quickly. Therefore, it becomes possible to avoid the failure promptly by measuring that and updating the weighting factor. It is particularly important that the load balancer autonomously performs such

a load-distribution function and a failure evasion function. That is, path control can be performed based only on the information acquired from the traffic which passes the load balancer, without acquiring any path information and failure information from other routers, servers, or the like.

[0070] Although the foregoing explanation of the example has been given of the case that a single provider uses the load balancer, a plurality of providers can use the respective load balancers in parallel. Furthermore, although the foregoing explanation has been given of the load-distribution processing being executed for a single destination as an example, it is also possible to simultaneously perform load-distribution processing on another destination while performing load-distribution processing to the arbitrary destination.

[0071] Furthermore, the target to which the load balancer distributes traffic is not limited to a path of the same kind. For example, even in a case that three paths which can reach the same destination exist, the first path uses the label switching path of the MPLS, the second path uses the virtual connection of the PPP, and the third path is connected to other routers using the VC of the ATM, load distribution among these paths is possible. The reason is because the path control method and apparatus according to the present invention are not dependent on specific means of each path, as long as the use bandwidth of the traffic which passes the load balancer can be measured with respect to a plurality of arbitrary paths.

[0072] The flow management method in the flow classifying unit 1 will be explained next with reference to Fig. 14. Fig. 14 is a flowchart of the flow management method. In the present embodiment, when the ratio of

weighting factors, W11:W12, changes, movement of the flow between paths occurs. When there is a difference in time delay between two paths at this time, the sequence of the packets belonging to the same flow can change on the way. For example, when the TCP or the like is used, changing of the packet sequence within the same flow can cause degradation of the performance.

[0073] As shown in Fig. 14, even when a weighting factor changes, the flow classifying unit 1 performs management in a way that packets belonging to the same flow can surely pass along the same path. In the method, a flow storage table 7 is provided, and the flow to which the passing packet belongs and the path assigned to the flow are registered in association with each other. First, when a packet is input, the flow storage table 7 is referred to check if a flow to which the packet belongs has already been registered (step 110). Consequently, when the flow has already been registered (step 111: YES), the packet is classified into the group corresponding to the flow (step 112). On the other hand, when the flow is not registered (step 111: NO), the packet is classified into the group whose weighting factor needs to be increased, and the corresponding flow and group are newly registered into the flow storage table 7 (step 113).

[0074] Fig. 15 is an example of the flow storage table 7 when a TCP connection is used as a flow. A flow is identified with four numbers, namely the sender address, the destination address, the sender port number, and the destination port number, as an identifier, and an output group is associated therewith. In the example, the packet whose flow identifier is {SIP1, DIP1, SPORT1, DPORT1} is classified into G11, and the packet whose flow identifier is {SIP2, DIP2, SPORT2, DPORT2} is classified into G12.

When the flow identifier of an input packet matches with any of them, the packet is classified into a corresponding output group. When the flow identifier of an input packet matches with none of them, the packet is classified into a group whose weighting factor is to be increased.

[0075] Fig. 16 is an example of a flow assignment table 8 which can be used at the time of determining into which group an unregistered flow is registered. In the flow assignment table 8, the current weighting factor (the current value of the registered number of flows) and a target weight (the target value of the registered number of flows) are associated with each other for each group. Assuming that the ratio of weighting factors W11:W12 is 1:2, a packet whose flow identifier is {SIP1, DIP1, SPORT3, DPORT1} is input, the corresponding flow identifier is not registered in the flow storage table 7 of Fig. 15. With reference to the flow assignment table 8, it is understood that the current weight of G12 is smaller than the target weight.

[0076] Therefore, this packet is registered into G12, a new flow identifier {SIP1, DIP1, SPORT3, DPORT1} is registered into the flow storage table 7 in association with the group G12, and the number of registered flows corresponding to the group G12 is increased by 1. For example, in the current Internet, the average sustaining time of a flow is several seconds, and sequential execution of the method can permit the number of flows proportional to the weighting factor to be output to each path, without changing the sequence of packets that belong to the same flow. An aging timer can be used to cancel registration of old flows in the flow storage table 7.

[0077]

[Effect of the Invention]

According to the present invention, as explained above, flows of input packets are classified into respective groups based on preset weighting factors thereof, the flows are output to different paths for the groups, that of bandwidths used in the paths which is used by each flow is measured group by group, and the weighting factor of each group is adjusted based on the ratio of the unit use bandwidths. Therefore, the bandwidths to be used by flows passing through the individual paths are averaged, thus making it possible to dynamically optimize the distribution of traffic over a plurality of paths in a network, so that network resources can be used efficiently.

[Brief Description of Drawings]

[Fig. 1] A schematic configuration diagram of a packet transfer system to which a path control method according to one embodiment of the present invention is adapted.

[Fig. 2] A concept diagram of a plurality of paths.

[Fig. 3] A block diagram of a configuration example of a load balancer.

[Fig. 4] A flowchart of a path control process in the load balancer.

[Fig. 5] An explanatory diagram of an example of an adjustment method for weighting factors.

[Fig. 6] An example of packet classification when a fourth layer connection of an OSI reference model is taken as a flow.

[Fig. 7] An example of packet classification when a third layer address of the OSI reference model is taken as a flow.

[Fig. 8] A configuration example of a flow classifying unit that classifies packets with a hash function.

[Fig. 9] An example of packet output when a path is distinguished with LSP of an MPLS network.

[Fig. 10] An example of packet output when a path is distinguished with a virtual connection of PPP.

[Fig. 11] An example of packet output when a path is distinguished with a virtual connection of an ATM network.

[Fig. 12] An example of packet output when a path is distinguished with a VLAN identifier of a VLAN network.

[Fig. 13] An example of packet output when a path is distinguished with an output physical interface.

[Fig. 14] A flowchart of a flow management method.

[Fig. 15] A configuration example of a flow storage table.

[Fig. 16] A configuration example of a flow assignment table.

[Explanations of Letters or Numerals]

N11 ··· Packet transfer network, B11 ··· Load balancer (Path control apparatus), R11, R12, R13 ··· Router, T11, T12, T13 ··· Terminal, P11, P12 ··· Path, G11, G12 ··· Group, W11, W12 ··· Weighting factor, BW11, BW12 ··· Unit use bandwidth, N12 ··· MPLS network, N13 ··· PPP network, N14 ··· ATM network, N15 ··· VLAN network, N16 ··· Network of physically divided path, 1 ··· Flow classifying unit, 2, 3 ··· Flow output unit, 4 ··· Bandwidth measuring unit, 5 ··· Unit bandwidth calculator, 6 ··· Weighting factor adjusting unit, 7 ··· Flow storage table, 8 ··· Flow assignment table.

[Fig. 1]

N11 Packet transfer network

T11, T12, T13 Terminal

R11, R12, R13 Router

B11 Load balancer

[Fig. 2]

T11, T12, T13 Terminal

B11 Load balancer

P11, P12 Path

R13 Router

[Fig. 3]

1 Flow classifying unit

2, 3 Flow output unit

4 Bandwidth measuring unit

5 Unit bandwidth calculator

6 Weighting factor adjusting unit

P11, P12 Path

G11, G12 Group

[Fig. 4]

100: Classify flows into groups based on weighting factors

101: Output flows to paths corresponding to groups

102: Measure bandwidth for each path

103: Calculate bandwidth per unit weighting factor

104: Adjust weighting factor

[Fig. 5]

Weighting factor for each group

Weighting factor W11 for group G11

Weighting factor W12 for group G12

When bandwidth per weighting factor is BW11>BW12, increase W11

When bandwidth per weighting factor is BW11=BW12, Hold W11, W12

When bandwidth per weighting factor is BW11<BW12, Increase W12

[Fig. 6]

Fourth layer connection

1 Flow classifying unit

Classify fourth layer connection of OSI reference model into groups based on weighting factors

G11, G12 Group

[Fig. 7]

Third layer address

1 Flow classifying unit

Assume a sequence of packets in which at least sender/destination address of third layer of OSI reference model is identical as a flow based on weighting factors, and classify the packets into groups

G11, G12 Group

[Fig. 8]

Input packet

Address, port number

21 Hash value calculator

Hash value

22 Hash value group classifying unit

Packet with hash value of 4 to 9

Packet with hash value of 1 to 3

To group of path P11

To group of path P12

[Fig. 9]

G11 Group

2 Flow output unit

Output packets affixed with label corresponding to path P11

MPLS network

[Fig. 10]

G11 Group

2 Flow output unit

Output packets affixed with identifier corresponding to path P11

Virtual connection

PPP network

[Fig. 11]

G11 Group

2 Flow output unit

Output packets affixed with identifier corresponding to path P11

Virtual connection or virtual path

ATM network

[Fig. 12]

Group G11

2 Flow output unit

Output packets affixed with VLAN identifier corresponding to path P11

VLAN network

[Fig. 13]

Group G11

2 Flow output unit

Output packets to node of next hop corresponding to path P11

Path

Network with a path branched by node of next hop

[Fig. 14]

7 Flow storage table

110 Search when flow to which packets belong has already been registered

111 Is flow registered?

112 Classify packets to registered group

113 Classify packets to group whose weighting factor is to be increased, and register flow

[Fig. 15]

7 Flow storage table

Flow identifier Group

Sender address

Destination address

Sender port number

Destination port number

[Fig. 16]

Flow assignment table

Group

Current weight (Current value of registered number of flows)

Target weight (Target value of registered number of flows)

(2)

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報**（Ａ）　(11)特許出願公開番号
特開2003－163687
（P2003－163687A）
(43)公開日　平成15年６月６日(2003.6.6)

(51)Int.Cl.'	識別記号	ＦＩ	テーマコード'(参考)
Ｈ０４Ｌ　12/56	１００	Ｈ０４Ｌ　12/56	１００Ｚ　５Ｋ０３０
	２００		２００Ｚ

審査請求　未請求　請求項の数30　ＯＬ　（全 14 頁）

(21)出願番号　特願2001－359254(P2001－359254)

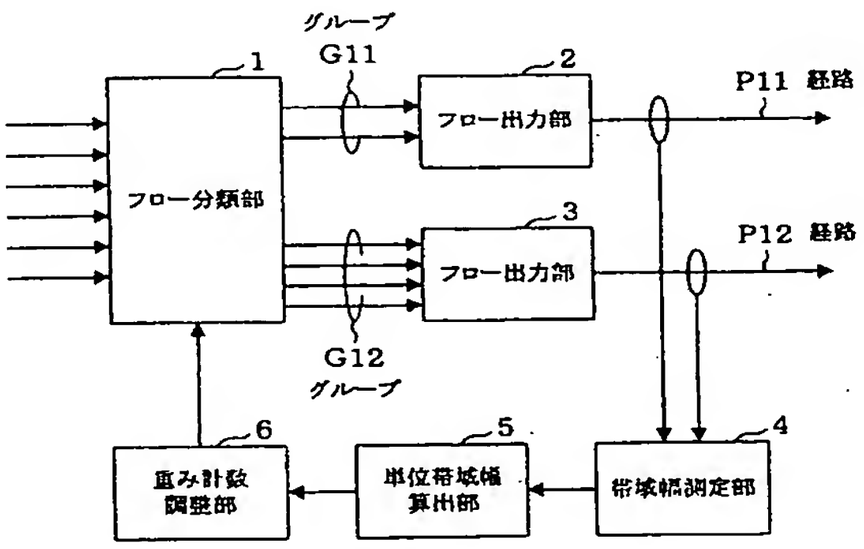
(22)出願日　平成13年11月26日(2001.11.26)

(71)出願人　000004226
日本電信電話株式会社
東京都千代田区大手町二丁目３番１号
(72)発明者　家永　憲人
東京都千代田区大手町二丁目３番１号　日本電信電話株式会社内
(72)発明者　宮本　正和
東京都千代田区大手町二丁目３番１号　日本電信電話株式会社内
(74)代理人　100064621
弁理士　山川　政樹

Ｆターム(参考)　5K030 GA01 HA08 LB05 LC01 LC11 LE03 MB09

(54)【発明の名称】　経路制御方法および装置

(57)【要約】
【課題】　ネットワーク内の複数経路にわたるトラフィックの配分を動的に最適化できるようにする。
【解決手段】　フロー分類部１では、入力されたパケットのフローを予め設定されている各グループの重み係数に基づきいずれかのグループへ分類し、フロー出力部２，３では、これらフローを当該グループごとに異なる経路へ出力する。帯域幅測定部４では、各経路で使用されている帯域幅のうち各フローで使用されている帯域幅をグループごとに測定し、単位帯域幅算出部５では、測定された各グループの使用帯域幅を当該グループに予め割り当てられている重み係数で除算することにより各グループごとに単位使用帯域幅を算出する。重み係数調節部６では、各グループの単位使用帯域幅に基づき各グループの重み係数を調節する。



1

【特許請求の範囲】

【請求項１】　パケット転送ネットワークに接続されて、入力されたパケットをその属性から当該パケットの属するフローを識別し、そのフローを単位として前記パケットをいずれかの経路へ出力する経路制御装置で用いる経路制御方法において、

入力されたパケットのフローを予め設定されている各グループの重み係数に基づきいずれかのグループへ分類するステップと、

前記各グループのフローを当該グループごとに異なる経路へ出力するステップと、

前記各経路で使用されている帯域幅を前記グループごとに測定するステップと、

測定された前記各帯域幅の比に基づき前記各グループの前記重み係数を調節するステップとを有することを特徴とする経路制御方法。

【請求項２】　請求項１記載の経路制御方法において、測定された前記各帯域幅を当該グループに予め割り当てられている重み係数で除算することにより前記各グループごとに単位使用帯域幅を算出するステップをさらに有し、

前記重み係数を調整する際、前記各単位使用帯域幅に基づき前記各グループの重み係数を調節することを特徴とする経路制御方法。

【請求項３】　請求項２記載の経路制御方法において、前記重み係数を調整する際、前記各グループの単位使用帯域幅が平均化されるようにそれぞれの重み係数を増減させることを特徴とする経路制御方法。

【請求項４】　請求項２記載の経路制御方法において、前記重み係数を調整する際、当該グループの単位使用帯域幅が他のグループの単位使用帯域幅よりも大きい場合は当該グループの重み係数を増加させ、他のグループの単位使用帯域幅よりも小さい場合は当該グループの重み係数を減少させることを特徴とする経路制御方法。

【請求項５】　請求項１記載の経路制御方法において、前記重み係数を調整する際、前記各グループの重み係数の比が前記各帯域幅の比と等しくなるように前記重み係数を調節することを特徴とする経路制御方法。

【請求項６】　請求項１記載の経路制御方法において、前記フローを分類する際、前記各グループに分類されるフローの数がそれぞれの重み係数に比例するように分類することを特徴とする経路制御方法。

【請求項７】　請求項１記載の経路制御方法において、前記フローを分類する際、前記各フローとして、ＯＳＩ参照モデル第４層のコネクションを用いることを特徴とする経路制御方法。

【請求項８】　請求項１記載の経路制御方法において、前記フローを分類する際、ＯＳＩ参照モデル第３層の送元または宛先アドレスの少なくとも一方が同一である一連の入力パケットを同一フローとすることを特徴とする

2

経路制御方法。

【請求項９】　請求項１記載の経路制御方法において、前記フローを分類する際、入力パケットの送元アドレス、宛先アドレス、送元ポート番号、および宛先ポート番号の少なくとも１つを引数に含むハッシュ関数により前記入力パケットのハッシュ値を計算し、得られた前記ハッシュ値が同一の入力パケットを同一グループに分類することを特徴とする経路制御方法。

【請求項１０】　請求項１記載の経路制御方法において、

前記フローを分類する際、分類した入力パケットの各フローを記憶するフロー記憶テーブルを用い、新たな入力パケットが前記フロー記憶テーブルに記憶されているいずれかのフローに属する場合は、当該入力パケットをそのフローと同じグループに分類し、前記入力パケットが前記フロー記憶テーブルに記憶されていない新たなフローに属する場合は、前記各グループのうち当該グループの重み係数を増加させるべきグループへ前記新たなフローを分類することを特徴とする経路制御方法。

【請求項１１】　請求項１記載の経路制御方法において、

前記経路として、ＭＰＬＳのＬＳＰに基づき区別される経路を用いることを特徴とする経路制御方法。

【請求項１２】　請求項１記載の経路制御方法において、

前記経路として、ＰＰＰの仮想コネクションに基づき区別される経路を用いることを特徴とする経路制御方法。

【請求項１３】　請求項１記載の経路制御方法において、

前記経路として、ＡＴＭの仮想コネクションＶＣまたは仮想パスＶＰに基づき区別される経路を用いることを特徴とする経路制御方法。

【請求項１４】　請求項１記載の経路制御方法において、

前記経路として、ＩＥＥＥ８０２．１Ｑに規定されるＶＬＡＮに基づき区別される経路を用いることを特徴とする経路制御方法。

【請求項１５】　請求項１記載の経路制御方法において、

前記経路として、当該経路制御装置の次のホップのノードに基づき区別される経路を用いることを特徴とする経路制御方法。

【請求項１６】　パケット転送ネットワークに接続されて、入力されたパケットをその属性から当該パケットの属するフローを識別し、そのフローを単位として前記パケットをいずれかの経路へ出力する経路制御装置経路制御装置において、

入力されたパケットのフローを予め設定されている各グループの重み係数に基づきいずれかのグループへ分類するフロー分類部と、

前記各グループのフローを当該グループごとに異なる経路へ出力するフロー出力部と、
前記各経路で使用されている帯域幅を前記グループごとに測定する帯域幅測定部と、
この帯域幅測定部で測定された前記各グループの帯域幅の比に基づき前記各グループの前記重み係数を調節する重み係数調節部とを備えることを特徴とする経路制御装置。

【請求項１７】　請求項１６記載の経路制御装置において、
前記帯域幅測定部で測定された前記各帯域幅を当該グループに予め割り当てられている重み係数で除算することにより前記各グループごとに単位使用帯域幅を算出する単位帯域幅算出部をさらに備え、
前記重み係数調整部は、この単位帯域幅算出部で算出された前記各単位使用帯域幅に基づき前記各グループの重み係数を調節することを特徴とする経路制御装置。

【請求項１８】　請求項１７記載の経路制御装置において、
前記重み係数調整部は、前記各グループの単位使用帯域幅が平均化されるようにそれぞれの重み係数を増減させることを特徴とする経路制御装置。

【請求項１９】　請求項１７記載の経路制御装置において、
前記重み係数調整部は、当該グループの単位使用帯域幅が他のグループの単位使用帯域幅よりも大きい場合は当該グループの重み係数を増加させ、他のグループの単位使用帯域幅よりも小さい場合は当該グループの重み係数を減少させることを特徴とする経路制御装置。

【請求項２０】　請求項１６記載の経路制御装置において、
前記重み係数調整部は、前記各グループの重み係数の比が前記各帯域幅の比と等しくなるように前記重み係数を調節することを特徴とする経路制御装置。

【請求項２１】　請求項１６記載の経路制御装置において、
前記フロー分類部は、前記各グループに分類されるフローの数がそれぞれの重み係数に比例するように分類することを特徴とする経路制御装置。

【請求項２２】　請求項１６記載の経路制御装置において、
前記フロー分類部は、前記各フローとして、ＯＳＩ参照モデル第４層のコネクションを用いることを特徴とする経路制御装置。

【請求項２３】　請求項１６記載の経路制御装置において、
前記フロー分類部は、ＯＳＩ参照モデル第３層の送元または宛先アドレスの少なくとも一方が同一である一連の入力パケットを同一フローとすることを特徴とする経路制御装置。

【請求項２４】　請求項１６記載の経路制御装置において、
前記フロー分類部は、入力パケットの送元アドレス、宛先アドレス、送元ポート番号、および宛先ポート番号の少なくとも１つを引数に含むハッシュ関数により前記入力パケットのハッシュ値を計算するハッシュ値計算部と、このハッシュ値計算部で得られたハッシュ値が同一の入力パケットを同一グループに分類するハッシュ値グループ分類部とを有することを特徴とする経路制御装置。

【請求項２５】　請求項１６記載の経路制御装置において、
前記フロー分類部は、分類した入力パケットの各フローを記憶するフロー記憶テーブルを有し、新たな入力パケットが前記フロー記憶テーブルに記憶されているいずれかのフローに属する場合は、当該入力パケットをそのフローと同じグループに分類し、前記入力パケットが前記フロー記憶テーブルに記憶されていない新たなフローに属する場合は、前記各グループのうち当該グループの重み係数を増加させるべきグループへ前記新たなフローを分類することを特徴とする経路制御装置。

【請求項２６】　請求項１６記載の経路制御装置において、
前記経路として、ＭＰＬＳのＬＳＰに基づき区別される経路を用いることを特徴とする経路制御装置。

【請求項２７】　請求項１６記載の経路制御装置において、
前記経路として、ＰＰＰの仮想コネクションに基づき区別される経路を用いることを特徴とする経路制御装置。

【請求項２８】　請求項１６記載の経路制御装置において、
前記経路として、ＡＴＭの仮想コネクションＶＣまたは仮想パスＶＰに基づき区別される経路を用いることを特徴とする経路制御装置。

【請求項２９】　請求項１６記載の経路制御装置において、
前記経路として、ＩＥＥＥ８０２．１Ｑに規定されるＶＬＡＮに基づき区別される経路を用いることを特徴とする経路制御装置。

【請求項３０】　請求項１６記載の経路制御装置において、
前記経路を識別する際、当該経路制御装置の次のホップのノードに基づき区別される経路を用いることを特徴とする経路制御装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、経路制御方法および装置に関し、特にパケット転送ネットワーク内の複数経路に対して動的に負荷を分散する経路制御方法および装置に関するものである。

【０００２】

【従来の技術】近年の通信環境は、アクセスラインの高速化、すなわちサービス加入者端末と最寄の加入者収容装置との間の伝送速度が飛躍的に向上したにもかかわらず、キャリアのコアネットワークの高速化は、さまざまな要因から遅れており、少数のユーザーのトラフィックによってネットワークのコアの帯域幅を簡単に使い切ってしまうこともあり得る。

【０００３】このような状況において、ネットワーク資源の利用効率を向上するために、近年、トラフィック・エンジニアリング（以下、ＴＥという：Traffic Engineering）の分野が精力的に研究されている。従来のＩＰネットワークでは、終点までの最もコストが低い経路（最短経路）に沿ったパケット転送が主流であったが、ＴＥにおいては第２、第３にコストが低い経路などもパケット転送に参加させることにより、特定の中継回線に負荷が集中することを回避し、これによってネットワーク資源の利用効率を向上することを目的としている。

【０００４】ＴＥでは、例えば、ユーザーの各拠点を結ぶ回線の必要帯域幅を、ネットワーク側での測定やユーザーからの要求によって見積もり、ネットワーク資源を最も効率良く利用できるように、回線に割り当てる経路の最適化計算を行う。しかし、この計算は複雑であり、大規模なネットワークでは最適な解を得るために長時間を要する。結果的に、トラフィックの増減に伴う経路の設定変更に数日以上を要することも希ではない。このように、長時間をかけてネットワーク資源を最適化していく方法を、以下では静的なＴＥと呼ぶことにする。

【０００５】先に述べたアクセスラインの高速化に伴って、インターネットのトラフィックとして、不特定多数を相手にしたバースト性の高いトラフィックが急増している。すなわち、従来の専用線や電話とは異なり、通信の相手が一定しておらず、しかも使用帯域幅がデータの転送時と非転送時とで大きな落差が生じる。このように経路や使用帯域幅が短時間で大きく変動するトラフィックに対しては、従来の静的なＴＥのみでは効率的なネットワーク資源の利用は困難であり、より短時間の変動に対して敏感に応答し、資源の最適化を行う方法が強く求められている。このような方法を、以下では動的なＴＥと呼ぶことにする。

【０００６】

【発明が解決しようとする課題】しかしながら、このような動的なＴＥでは従来の負荷分散方法を適用できないという問題点があった。次に、従来の負荷分散方法を動的なＴＥへ適用する際の技術的な課題を述べる。一般にＴＥを行うためには、①ネットワークの各経路の負荷すなわち利用度合を測定し、②測定結果を分析して資源の割り当てを最適化する計算を行い、③計算結果に基づいた新しい経路によってトラフィックを転送する、という三段階を経る。

【０００７】ＩＰネットワークにおける①の負荷の測定は、負荷情報を監視するサーバーから監視対象である各経路制御装置に対して、ＳＮＭＰ（Simple Network Management Protocol）を利用して、遠隔で情報を取得するのが一般的である。しかし、この方法を動的なＴＥに適用するには課題が多い。以下ではＳＮＭＰを用いた場合を例として負荷情報を収集する際の課題点を述べる。第１に、過負荷リンク、シェーパーの扱いが困難である点が挙げられる。過負荷になったリンク（物理速度を使い切ってしまったリンク）、あるいはシェーパー（一定の帯域幅にトラフィックを整形する装置）が経路の途中に存在する場合、負荷情報として単にリンクの流量を取得するだけでは情報不足である。すなわち、最大速度を使い切った状態では、本当はいくら超過しているのか、どれだけのトラフィックを他の経路に振り分ければよいのかという情報が得られない。

【０００８】この問題を厳密に解決するためには、各経路制御装置の入力帯域幅と出力帯域幅の差分を、各経路に対して測定する必要がある。しかも、経路制御装置は一般的に入力インターフェイスおよび出力インターフェイスが複数あり、トラフィックは複数の経路制御装置内で離合集散を繰り返すため、あるリンクの輻輳の原因を調べるには、すべての経路を通って合流してくるトラフィック量や、経路制御装置での経路制御表などまで調べる必要がある。しかし、これらの情報を高速に収集し、解析するのは非常に難しい。したがって、過負荷になったリンクや、シェーパーが経路の途中に存在する場合でも、負荷の度合や他の経路へ分散が必要な量を見積もるための、簡単かつ効果的な方法が必要である。

【０００９】第２に、サーバー設置により、新たな複雑性が増加する点が挙げられる。動的な負荷分散では、数秒周期で負荷情報を時々刻々収集する必要がある。大規模なネットワークで、頻繁に負荷情報を収集するには、各経路制御装置が個別に測定や分析などを行うのではなく、負荷情報の収集を行う専用のサーバーが、一括して情報を集め分析する方が、効率がよい。しかしこれは、ネットワークの構築に新たな複雑性を持ち込むことになる。例えば、負荷情報用の別回線の設定や二重化、さらにはサーバーの二重化などを考慮しなければならない。このようなサーバーを使用せず、しかも効率のよい情報収集の方法が、強く求められている。

【００１０】第３に、管理対象数の増大が挙げられる。各経路を適切に迂回させるためには、経路毎の負荷をネットワーク内のすべての経路制御装置に対して測定する必要がある。しかし、大規模なネットワークでは経路の数や経路制御装置の数が劇的に増大するため、これらの経路の各経由地点に対して測定を行い、またその結果を解析するのは難しい。結果的に、負荷の変動に対する最適化計算に長時間を要することとなり、動的なＴＥに適用するのは困難である。少ない管理対象で効果的に負荷

分散を行う方法が必要である。

【００１１】第４に、故障時の対応が困難な点が挙げられる。動的なＴＥではネットワークの障害に瞬時に対応する能力が求められる。そのため、ある経路のトラフィック量が少ない場合、その原因が、リンクの故障や経路制御の障害によるものなのか、あるいは単に空いているからだけなのかを、短時間で切り分ける必要がある。空いているのであればそこにトラフィックを積極的に流すのが好ましいであろうし、逆に障害が発生しているのであれば、その経路は利用しないようにするのが望ましいからである。

【００１２】しかし、実際のネットワークでは、ＳＮＭＰのみを使用してこのような障害を切り分けることは困難である。例えば経路上にＳＮＭＰの管理機能を持たないリピータやスイッチが存在する場合、その部分で生じた障害を検出することができない。また、障害の原因はインターフェイスの故障だけではなく、経路制御の過渡的な状態におけるループの発生など、さまざまな要素が考えられる。したがって、経路制御上の障害を正確に検出するには、このようなリピータやスイッチを含めたパケット転送にかかわるすべての要素の状態を監視することに加えて、各経路制御装置が保持している経路制御情報（経路制御プロトコルが交換している情報）や、経路制御表を取得する必要がある。

【００１３】さらに問題を難しくしているのは、同一のネットワーク内でも複数の経路制御プロトコルが併用されることが多い点である。すなわち、手動操作による静的な経路の設定や、ＯＳＰＦ（Open Shortest Path First）やＲＩＰ（Routing Information Protocol）などの動的な経路制御プロトコルを利用した経路の設定が、併用されるのが一般的であるため、経路制御の障害を検出するには、これらの静的あるいは動的な経路の設定すべてに対して、各々に最適な方法で情報を収集し、それらを相互に変換して統一的に解析する必要がある。例えば、ＯＳＰＦに対してはリンク状態を表す情報を取得し、ＲＩＰでは隣接経路制御装置に対するアップ・ダウン情報や、どの隣接装置から経路を学習したかなどを調べる必要がある。したがって、ネットワークに障害が発生した場合においても、速やかにそれを検出して経路の最適化を行える方法が必要である。

【００１４】このように、従来の負荷分散方法を動的なＴＥへ適用するには多くの課題があり、実用的なレベルでネットワーク内の経路を動的に最適化することは困難であった。本発明はこのような課題を解決するためのものであり、ネットワーク内の複数経路にわたるトラフィックの配分を動的に最適化することができる経路制御方法および装置を提供することを目的としている。

【００１５】

【課題を解決するための手段】このような目的を達成するために、本発明にかかる経路制御方法は、パケット転送ネットワークに接続されて、入力されたパケットをその属性から当該パケットの属するフローを識別し、そのフローを単位としてパケットをいずれかの経路へ出力する経路制御装置で用いる経路制御方法において、入力されたパケットのフローを予め設定されている各グループの重み係数に基づきいずれかのグループへ分類するステップと、各グループのフローを当該グループごとに異なる経路へ出力するステップと、　各経路で使用されている帯域幅をグループごとに測定するステップと、測定された各帯域幅の比に基づき各グループの重み係数を調節するステップとを有するものである。

【００１６】さらに、測定された各帯域幅を当該グループに予め割り当てられている重み係数で除算することにより各グループごとに単位使用帯域幅を算出するステップを設け、重み係数を調整する際、各単位使用帯域幅に基づき各グループの重み係数を調節するようにしてもよい。

【００１７】重み係数を調整する際、各グループの単位使用帯域幅が平均化されるようにそれぞれの重み係数を増減させるようにしてもよい。この他、当該グループの単位使用帯域幅が他のグループの単位使用帯域幅よりも大きい場合は当該グループの重み係数を増加させ、他のグループの単位使用帯域幅よりも小さい場合は当該グループの重み係数を減少させるようにしてもよい。また、重み係数を調整する際、各グループの重み係数の比が各帯域幅の比と等しくなるように重み係数を調節するようにしてもよい。

【００１８】フローを分類する際、各グループに分類されるフローの数がそれぞれの重み係数に比例するように分類するようにしてもよい。この他、各フローとして、ＯＳＩ参照モデル第４層のコネクションを用いてもよく、あるいは、ＯＳＩ参照モデル第３層の送元または宛先アドレスの少なくとも一方が同一である一連の入力パケットを同一フローとするようにしてもよい。

【００１９】また、フローを分類する際、入力パケットの送元アドレス、宛先アドレス、送元ポート番号、および宛先ポート番号の少なくとも１つを引数に含むハッシュ関数により入力パケットのハッシュ値を計算し、得られたハッシュ値が同一の入力パケットを同一グループに分類するようにしてもよく、あるいは、分類した入力パケットの各フローを記憶するフロー記憶テーブルを用い、新たな入力パケットがフロー記憶テーブルに記憶されているいずれかのフローに属する場合は、当該入力パケットをそのフローと同じグループに分類し、入力パケットがフロー記憶テーブルに記憶されていない新たなフローに属する場合は、各グループのうち当該グループの重み係数を増加させるべきグループへ新たなフローを分類するようにしてもよい。

【００２０】経路としては、ＭＰＬＳのＬＳＰに基づき区別される経路を用いるようにしてもよい。この他、Ｐ

ＰＰの仮想コネクション、ＡＴＭの仮想コネクションＶＣまたは仮想パスＶＰ、ＩＥＥＥ８０２．１Ｑに規定されるＶＬＡＮ、あるいは当該経路制御装置の次のホップのノードに基づき、それぞれ区別される経路を用いるしてもよい。

【００２１】また、本発明にかかる経路制御装置は、パケット転送ネットワークに接続されて、入力されたパケットをその属性から当該パケットの属するフローを識別し、そのフローを単位としてパケットをいずれかの経路へ出力する経路制御装置経路制御装置において、入力されたパケットのフローを予め設定されている各グループの重み係数に基づきいずれかのグループへ分類するフロー分類部と、各グループのフローを当該グループごとに異なる経路へ出力するフロー出力部と、各経路で使用されている帯域幅をグループごとに測定する帯域幅測定部と、この帯域幅測定部で測定された各グループの帯域幅の比に基づき各グループの重み係数を調節する重み係数調節部とを備えるものである。

【００２２】さらに、帯域幅測定部で測定された各帯域幅を当該グループに予め割り当てられている重み係数で除算することにより各グループごとに単位使用帯域幅を算出する単位帯域幅算出部を設け、重み係数調整部では、この単位帯域幅算出部で算出された各単位使用帯域幅に基づき各グループの重み係数を調節するようにしてもよい。

【００２３】重み係数調整部では、重み係数を調整する際、各グループの単位使用帯域幅が平均化されるようにそれぞれの重み係数を増減させるようにしてもよい。この他、当該グループの単位使用帯域幅が他のグループの単位使用帯域幅よりも大きい場合は当該グループの重み係数を増加させ、他のグループの単位使用帯域幅よりも小さい場合は当該グループの重み係数を減少させるようにしてもよい。また、重み係数を調整する際、各グループの重み係数の比が各帯域幅の比と等しくなるように重み係数を調節するようにしてもよい。

【００２４】フロー分類部では、フローを分類する際、各グループに分類されるフローの数がそれぞれの重み係数に比例するように分類するようにしてもよい。この他、各フローとして、ＯＳＩ参照モデル第４層のコネクションを用いるようにしてもよく、あるいはＯＳＩ参照モデル第３層の送元または宛先アドレスの少なくとも一方が同一である一連の入力パケットを同一フローとするようにしてもよい。

【００２５】また、フロー分類部でフローを分類する際、ハッシュ値計算部により、入力パケットの送元アドレス、宛先アドレス、送元ポート番号、および宛先ポート番号の少なくとも１つを引数に含むハッシュ関数により入力パケットのハッシュ値を計算し、ハッシュ値グループ分類部により、ハッシュ値計算部で得られたハッシュ値が同一の入力パケットを同一グループに分類するようにしてもよい。

【００２６】フロー分類部では、フローを分類する際、分類した入力パケットの各フローを記憶するフロー記憶テーブルを設け、新たな入力パケットがフロー記憶テーブルに記憶されているいずれかのフローに属する場合は、当該入力パケットをそのフローと同じグループに分類し、入力パケットがフロー記憶テーブルに記憶されていない新たなフローに属する場合は、各グループのうち当該グループの重み係数を増加させるべきグループへ新たなフローを分類するようにしてもよい。

【００２７】経路として、ＭＰＬＳのＬＳＰに基づき区別される経路を用いるようにしてもよい。この他、ＰＰＰの仮想コネクション、ＡＴＭの仮想コネクションＶＣまたは仮想パスＶＰ、ＩＥＥＥ８０２．１Ｑに規定されるＶＬＡＮ、あるいは当該経路制御装置の次のホップのノードに基づき区別される経路を用いるようにしてもよい。

【００２８】

【発明の実施の形態】次に、本発明の実施の形態について図面を参照して説明する。図１は本発明の一実施の形態にかかる経路制御方法が適用されるパケット転送ネットワークの概略構成図である。パケット転送ネットワークＮ１１には、本発明の一実施の形態にかかる経路制御方法を用いた経路制御装置を有する１つの負荷分散装置Ｂ１１と３つのルーターＲ１１，Ｒ１２，Ｒ１３が設けられており、伝送手段Ｌ１１によって相互に接続されている。ルーターＲ１１，Ｒ１２，Ｒ１３は、パケットを転送する装置全般のことであり、スイッチ、ハブ、リピータ、パケット交換機などをこれらルーターとして利用することができる。以下では３つのルーターを用いた場合を例としているが、任意の数のルーターを用いることができる。

【００２９】負荷分散装置Ｂ１１には、伝送手段Ｌ１１を介して複数の端末Ｔ１１，Ｔ１２が接続されている。以下では１つの負荷分散装置Ｂ１１を用いた場合を例としているが、任意の数の負荷分散装置を用いることができる。また、負荷分散装置Ｂ１１と端末Ｔ１１，Ｔ１２とが伝送手段Ｌ１１により直接接続されているが、両者の間に任意の数のルーターやネットワークが存在していてもよい。さらに負荷分散装置Ｂ１１は、図１のようにネットワークの入口に設ける必要はなく、例えば端末Ｔ１１，Ｔ１２に内蔵したり、ルーターＲ１１，Ｒ１２に内蔵することも可能である。

【００３０】各伝送手段Ｌ１１は、すべて同一である必要はなく、例えばＴ１１－Ｂ１１間を接続する伝送手段と、Ｂ１１－Ｒ１１間を接続する伝送手段とで異なる技術を使用することができる。このほか、図には示していないが、パケット転送ネットワークＮ１１には他の負荷分散装置やルーターを経由して非常に多くの端末が接続されていてもよく、ネットワーク内Ｎ１１内の各部分を

流れるトラフィックの量は、時々刻々変化してもよい。

【００３１】図２は、端末Ｔ１１や端末Ｔ１２を送元として、負荷分散装置Ｂ１１を経由して端末Ｔ１３にパケットを転送する場合の複数経路を示す概念図である。前述した図１において、負荷分散装置Ｂ１１からルーターＲ１１，Ｒ１２，Ｒ１３を経由して、端末Ｔ１３へ向かう経路は複数考えられる。ここでは、理解を容易とするため、２つの経路Ｐ１１，Ｐ１２を使用する場合を考える。経路Ｐ１１として、負荷分散装置Ｂ１１→ルーターＲ１１→ルーターＲ１３→端末Ｔ１３なる経路を使用し、経路Ｐ１２として、負荷分散装置Ｂ１１→ルーターＲ１２→ルーターＲ１３→端末Ｔ１３なる経路を使用する。

【００３２】端末Ｔ１１，Ｔ１２から出力されたパケットは、他の端末からのトラフィックと離合集散を繰り返しながら、経路Ｐ１１と経路Ｐ１２のいずれかを通って端末Ｔ１３に到達する。経路途中の各ルーターＲ１１，Ｒ１２や、伝送手段Ｌ１１の負荷は時々刻々変化しているため、負荷分散装置Ｂ１１の目標は、各経路の負荷に応じて最もネットワーク資源の利用効率が高くなるようにこれらの経路Ｐ１１，Ｐ１２にトラフィックを転送することである。

【００３３】図３は負荷分散装置Ｂ１１の構成例を示すブロック図である。負荷分散装置Ｂ１１には、フロー分類部１、フロー出力部２，３、帯域幅測定部４、単位帯域幅算出部５および重み係数調整部６が設けられている。この負荷分散装置Ｂ１１では、入力されるパケットの属性、例えば送元アドレス、宛先アドレス、送元ポート番号、宛先ポート番号などを単独であるいは組み合わせて用いた情報に基づいて、同一属性を持つ各パケットを１つのフローとして管理している。また、これらフローを複数のグループに分類し、これらグループで使用される各経路の帯域幅を調整することにより、各経路の負荷を分散制御している。

【００３４】フロー分類部１は、各グループごとに予め設定されている重み係数に基づき、入力パケットが属するフローを各グループへ分類する機能部である。フロー出力部２，３は、各経路Ｐ１１，Ｐ１２ごとに設けられ、フロー分類部１で各グループに分類されたフローをそのグループ単位で当該経路へ出力する機能部である。帯域幅測定部４は、各経路で当該グループが使用している帯域幅を測定する機能部である。

【００３５】単位帯域幅算出部５は、帯域幅測定部４により測定された各グループの使用帯域幅を当該グループの重み係数で除算して、各グループの単位重み係数当たりの単位使用帯域幅を算出する機能部である。重み係数調整部６は、単位帯域幅算出部５で算出された各グループの単位使用帯域幅が平均化するよう、各グループの重み係数を調整する機能部である。これら機能部は、それぞれ個別に、ＣＰＵなどのマイクロプロセッサとその周辺回路からなるハードウェアと上記マイクロプロセッサで実行されるソフトウェアとを協働させて構成してもよく、ハードウェア回路のみで構成してもよい。

【００３６】以下では、各フローを２つのグループＧ１１，Ｇ１２へ分類するものとし、グループＧ１１のフローはフロー出力部２から経路Ｐ１１へ出力され、グループＧ１２のフローはフロー出力部３から経路Ｐ１２へ出力される場合を例として説明する。

【００３７】図４は負荷分散装置での経路制御処理を示すフローチャートである。端末Ｔ１１，Ｔ１２から入力されたパケットは、フロー分類部１において、各グループの重み係数に基づきグループＧ１１，Ｇ１２に分類される（ステップ１００）。分類の方法としては、例えば各グループに分類されるフローの数が重み係数に比例するように分類するものとする。図３では、入力フロー数が６で、グループＧ１１，Ｇ１２の重み係数Ｗ１１，Ｗ１２がそれぞれ１と２の場合が例として示されている。上記の分類方法によれば、入力フローのうちの２つはグループＧ１１へ、あとの４つはグループＧ１２へ分類されることになる。

【００３８】グループＧ１１，Ｇ１２はそれぞれフロー出力部２，３により、経路Ｐ１１，Ｐ１２へ出力される（ステップ１０１）。各経路Ｐ１１，Ｐ１２に出力されたパケットの流れは、帯域幅測定部４により、各グループごとにその使用帯域幅が逐次測定される（ステップ１０２）。この測定した結果に基づき、単位帯域幅算出部５では、単位重み係数当たりの帯域幅を計算する（ステップ１０３）。例えば、測定された経路Ｐ１１，Ｐ１２での使用帯域幅がそれぞれ１０と１２であった場合、これらをそれぞれ対応する重み係数１と２で除算すると、単位重み係数当たり単位使用帯域幅はそれぞれ１０と６になる。

【００３９】単位帯域幅算出部５で算出された各グループごとの単位使用帯域幅に基づき、重み係数調整部６で各グループの重み係数が調節される（ステップ１０４）。したがって、その後、フロー分類部１において各フローを再分類する際あるいは新たなフローを分類する際、ステップ１００へ戻って、重み係数調整部６により調整された重み係数が用いられることになり、各経路を通るフローで使用される帯域幅が平均化されることから、ネットワーク内の複数経路にわたるトラフィックの配分が動的に最適化されることになる。これにより、バースト性の高いトラフィックに対する、ネットワーク資源の利用効率が向上し、多数の顧客に低価格なパケット転送機能を提供することができるようになる。

【００４０】本発明は、ネットワークの負荷状況に応じて動的に転送方路を変更するパケット転送方法を提供するというひとつの側面を持っている。同一のＴＣＰコネクションに属するパケットをフローとみなす場合を例にして、具体的に説明すると、各グループに含まれるＴＣ

Ｐコネクション数の比が、各グループに割り当てられた重み係数の比に等しくなるようにトラフィックを分類して、それぞれ異なる経路に送出し、各経路に送出したグループの使用帯域幅をグループごとに測定し、測定した使用帯域幅を当該グループの重み係数で除すことにより単位重み係数当りの単位使用帯域幅をグループごと計算し、この単位重み係数当りの単位使用帯域幅がグループ間で均等になるように重み係数を最適化している。これにより、経路上に物理帯域幅一杯まで使い切っているリンクやシェーパーが存在する場合においても、各経路を通るＴＣＰセッションのスループットを定量的に比較し、最適化することができるため、前述した第１の課題を解決することができる。

【００４１】また、上記の測定や最適化の計算は、各負荷分散装置が他の負荷分散装置やルーターからの、負荷や障害に関する情報を使用せずに独立して行うことができるため、前述した第２の課題であるサーバーを設置する必要がない。また、上述の各グループの使用帯域幅は、各負荷分散装置が当該装置を通過するフローに対してのみ内部的に測定し処理することができるため、ネットワーク全体にわたって全経路の全通過点に対する情報収集は不要となる。したがって従来の方法に比較して測定対象数が少なくなるとともに、さらに測定処理が複数の負荷分散装置に分散されるため、前述した第３の課題を解決することができる。

【００４２】さらに、上述の方法により前述した第４の課題も解決される。これは、一般にＴＣＰなどのトランスポート層プロトコルが、輻輳制御機能を具備していることを利用している。例えば、ＴＣＰを使って通信する端末はネットワークの品質に応じて動的にコネクションの使用帯域幅を増減する。すなわち各ＴＣＰコネクションは、通信経路に障害や輻輳が発生した場合には、自動的に送信速度を低下させ、逆に通信経路が空いている場合には、自動的に送信速度を増加させる。

【００４３】本発明では、従来のように各経路の使用帯域幅を測定するのではなく、単位重み係数当りの使用帯域幅すなわち単位使用帯域幅を計算しており、これはＴＣＰコネクション当りの使用帯域幅に比例する。したがって、この単位使用帯域幅を複数の経路の間で比較することにより、単位使用帯域幅が極端に小さければ、何らかの障害が発生したと考えられるし、単位使用帯域幅が増加していれば、経路が空いていることがわかる。これにより、ある経路のトラフィックの減少がリンクの障害によるものか、単に空いているからだけなのかを把握でき、短時間で切り分けることが可能となる。

【００４４】なお、以上の説明では、帯域幅測定部４において、フロー出力部２，３からの出力に基づき使用帯域幅を測定しているが、グループＧ１１，Ｇ１２は、経路Ｐ１１，Ｐ１２と１対１で対応しているため、フロー分類部１からの出力で使用帯域幅を測定することもでき

る。また図４では、経路制御処理を一連の流れ処理として説明したが、経路制御処理における各ステップは、それぞれ単独で実行してもよい。特に、帯域幅測定部４での測定間隔は、周期的に行うこともできるし、意図的に変動させることもできる。あるいは必要に応じて行うようにしてもよい。

【００４５】次に、図５を参照して、重み係数調整部６における重み係数の調節方法について説明する。図５は重み係数の調整例を示す説明図である。グループＧ１１，Ｇ１２の重み係数をＷ１１，Ｗ１２とし、単位重み係数当たりの単位使用帯域幅をＢＷ１１，ＢＷ１２とすると、重み係数調整部６では、単位使用帯域幅ＢＷ１１，ＢＷ１２の大小に応じて重み係数Ｗ１１，Ｗ１２が増減される。すなわち、ＢＷ１１＞ＢＷ１２の場合には、Ｗ１１を増加させてＷ１２を減少させ、ＢＷ１１＜ＢＷ１２の場合には、Ｗ１１を減少させてＷ１２を増加させる。ＢＷ１１とＢＷ１２が同じ場合には、Ｗ１１，Ｗ１２の値を変更する必要はない。

【００４６】重み係数Ｗ１１，Ｗ１２を、各グループＧ１１，Ｇ１２に含まれるフローの数に比例するように調整した場合、結果的に経路Ｐ１１を経由するフローの数は増加し、経路Ｐ１２を経由するフローの数は減少する。したがって、フロー当たりの帯域幅が均等になる方向に系が変化する。なお、重み係数の調整方法については、単位使用帯域幅ＢＷ１１とＢＷ１２の絶対値や差の大小により、重み係数Ｗ１１とＷ１２の調節の度合いを変更することができる。

【００４７】以上では、各経路に対して単位重み係数当たりの単位使用帯域幅という概念を導入した。しかし、前述した重み係数を調節するステップ１０４（図４参照）において、各経路の単位使用帯域幅を均等にするような制御を行う場合には、前述の単位重み係数当たりの単位使用帯域幅を計算するステップ１０３を省略する構成が可能である。この場合、ステップ１０４では、ステップ１０２で測定された各経路の使用帯域幅を入力変数とし、各経路に対する重み係数の比が各経路の使用帯域幅の比と等しくなるように、各重み係数を調節すればよい。

【００４８】また、このような構成を用いる場合には、図３における帯域幅算出部５も同様に省略可能であり、このときの重み係数調整部６では、帯域幅測定部４で測定された各経路の使用帯域幅を入力とし、各経路の使用帯域幅の比に等しい重み係数を出力する。具体例で説明すると、例えば測定された経路Ｐ１１，Ｐ１２での使用帯域幅がそれぞれ１０と１２であった場合、コネクション当たりの帯域を均等にするためには、各経路Ｐ１１，Ｐ１２に対して１０：１２の比でコネクションを分類すればよい。

【００４９】ただし、このようにして経路の間でコネクションを移動させたことが原因で、Ｐ１１，Ｐ１２での

使用帯域幅が当初の１０：１２から変化したり、系が不安定になる可能性がある。したがって、重み係数の調節は、逐次得られる使用帯域幅に即して複数回に分けて徐々に変化させてもよい。あるいは、収束速度を加速する目的で、重み係数変更後の帯域の変動まで考慮して、重み係数の変動量を上記より意図的に大きくしてもよい。

【００５０】次に、図６，７を参照して、フロー分類部１におけるフローの分類方法について説明する。図６，７にフロー分類方法の一例を示す。図６のフロー分類方法では、ＯＳＩ参照モデルの第４層コネクションをフローとみなしている。ここでは、分類のための属性として、第４層プロトコルの例としてＴＣＰ（Transmission Control Protocol）を使用している。この場合、ＴＣＰの１コネクションが１フローに対応する。ＴＣＰは輻輳制御機能やフロー制御機能を備えており、データ量や遅延、パケット損失などの外部条件が同じであれば、各コネクションの帯域幅は平均化する傾向がある。したがって、上述のようにＢＷ１１とＢＷ１２を平均化するようにＷ１１，Ｗ１２を調節すれば、両経路Ｐ１１，Ｐ１２を経由するＴＣＰコネクションのすべての帯域幅が平均化する方向に系が変化することになる。これにより、ネットワーク全体での資源利用効率が向上する。

【００５１】図７のフロー分類方法では、ＯＳＩ参照モデルの第３層の送元または宛先アドレスの少なくとも一方が同一である一連のパケットをフローとみなしている。ここでは、分類のための属性として、第３層の例としてＩＰを使用し、送元ＩＰアドレスに基づいてフローを識別している。この場合、上述の第４層プロトコルによってフローを識別する方法に比較して、フローを識別するための負担を軽くすることができる。なぜなら、ＴＣＰコネクションを識別するためには、一般に送元ＩＰアドレス、宛先ＩＰアドレス、送元ポート番号、宛先ポート番号の４つの数字が必要であるが、図７の場合には、送元ＩＰアドレスのみでよいからである。この例では、送元アドレス当たりの帯域幅が、２つの経路で均等になるように系が変化することになる。ここで例えば、第４層プロトコルとしてＴＣＰを利用しており、さらに各送元ＩＰアドレスが平均して同じ程度のＴＣＰコネクションを使用している場合には、各送元ＩＰアドレスの端末が利用する帯域幅が均等になる。

【００５２】図６や図７で示した方法では、負荷分散装置を通過するパケットの属性に基づきどのフローに属するかをすべて検査し、記憶する必要がある。しかしながら本実施の形態において本質的に必要なのは、重み係数の比、Ｗ１１：Ｗ１２に比例した割合で入力フローをＧ１１とＧ１２とに分類することである。この目的のためには、ハッシュ関数を利用することができる。図８では、送元アドレス、宛先アドレス、送元ポート番号、宛先ポート番号を引数とし、１～９の整数を値域とするハッシュ関数を利用するフロー分類部１の構成例が示されており、ハッシュ値計算部２１とハッシュ値グループ分類部２２とが設けられている。

【００５３】ハッシュ関数としては、ハッシュ関数の値域の各微小部分に含まれるフローの数が均等になるものを採用することができる。このようなハッシュ関数の例として、例えば「（送元アドレス＋宛先アドレス＋送元ポート番号＋宛先ポート番号）ｍｏｄ９＋１」なる関数が挙げられる（但し、ＡｍｏｄＢはＡをＢで割り算した場合の余りを示す）。ハッシュ値計算部２１で各パケットに対して上記ハッシュ関数を計算すると、ハッシュ値として１～９の整数が得られる。インターネットのように十分フロー数が多い場合には、各ハッシュ値をとるフローの数はおよそ均等になる。なお、フロー分類の割合を細かく調整したい場合には、上記関数の分割数９を大きくすればよい。ハッシュ値グループ分類部２２では、この値域を先のＷ１１：Ｗ１２の比に分割し、それぞれグループＧ１１，Ｇ１２と対応付ける。これにより、結果的に入力フローを任意の割合で分類することができる。この方法の利点は、負荷分散装置を通過するフローを記憶する必要がないことである。

【００５４】次に、経路の区別について説明する。本実施の形態では、パケット転送ネットワークＮ１１の内部で、同一の宛先に対して複数の経路を設定する必要がある。従来のインターネットにおいては、同一の宛先に対するパケットは同一の経路を通るため、何らかの方法で経路を区別する方法を提供しなければならない。この目的のために利用することができる方法の例として、ＭＰＬＳ（Multi Protocol Label Switching）のＬＳＰ（Label Switched Path）、ＰＰＰ（Point-to-Point Protocol）の仮想コネクション、フレームリレーの仮想コネクション、ＡＴＭ（Asynchronous Transfer Mode）の仮想パスＶＰ（Virtual Path）や仮想チャネルＶＣ（Virtual Connection）、ＩＥＥＥ８０２．１ＱのＶＬＡＮ（Virtual Bridged Local Area Networks）、次のホップのノード（物理インターフェイス）、無線のチャネルや周波数、ＷＤＭ（Wavelength Division Multiplexing）における波長などに基づき経路を区別することができるが、複数の経路を区別できるのであれば、上記の方法に限定するものではない。

【００５５】図９では、ＭＰＬＳネットワークのＬＳＰで経路を区別する場合を示している。ＭＰＬＳのネットワークは、一般に、ラベルエッジルーター（Label Edge Router/LER）と、ラベルスイッチルーター（Label Switching Router/LSR）とが、相互に接続されて構成されている。ここで、ラベルエッジルーターは、ＭＰＬＳネットワークの外周部に設置され、外部のネットワークまたは端末からパケットを受信してＭＰＬＳネットワーク内にパケットを転送したり、ＭＰＬＳネットワーク内部からの受信したパケットを外部のネットワークや端末へ転送する働きを持つ。一方、ラベルスイッチルーター

は、ＭＰＬＳネットワーク内部に位置し、ラベルエッジルーターや他のラベルスイッチルーターから受信したパケットを、他のラベルエッジルーターやラベルスイッチルーターへ転送する働きを持つ。

【００５６】ＭＰＬＳネットワークにおいてパケットが転送される経路であるＬＳＰは、ＭＰＬＳネットワーク入口のラベルエッジルーター（Label Edge Router/LER）を始点とし、ＭＰＬＳネットワーク出口のラベルエッジルーターを終点とする、一連のラベルエッジルーターおよびラベルスイッチルーターに沿って設定される。パケット転送は、通常、ラベルエッジルーターにおいて、パケットの入力インターフェイスや宛先ＩＰアドレスなどの情報を取得し、これをもとに経路制御表を検索して、パケットを各転送クラス（Forwarding Equivalence Class/FEC）に分類する。

【００５７】次に、各転送クラスに対応するラベルの値、パケットを出力すべきインターフェイス、次ホップのラベルスイッチルーターのＯＳＩ参照モデル第２層アドレスなどを検索し、これらの情報に基づいてパケットをＭＰＬＳヘッダおよび第２層ヘッダでカプセル化した後、対応する次ホップのラベルスイッチルーターへ転送する。以上のようなＭＰＬＳネットワークにおいて、本発明を用いて負荷分散を行うにあたって、例えばＭＰＬＳネットワークの入口に本発明にかかる経路制御装置を有する負荷分散装置を設置し、負荷分散を行いたい転送クラスに対して、当該装置を始点とする複数のラベルスイッチパスを設定すればよい。これにより、任意の転送クラスに分類されたパケットを、各ラベルスイッチパスに対応するグループへそれぞれの重み係数に基づき分類して転送することができる。

【００５８】上記では、１つの転送クラスに対して複数のラベルスイッチパスを対応付ける例について説明したが、１つの転送クラスに対して１つのラベルスイッチパスを対応付けることも可能である。この場合は、各グループと転送クラスとが１対１で対応付けられ、入力パケットを転送クラスに分類するステップにおいて、重み係数に基づくグループへの分類も同時に行うことができる。

【００５９】なお、上記負荷分散装置を設置する場所は、ＭＰＬＳネットワークの入口部分に限るものではない。すなわち、任意の入口ラベルエッジルーターに始まり出口ラベルエッジルーターで終わるラベルスイッチパスの途中に上記負荷分散装置を設置し、この負荷分散装置を始点として上記出口ラベルエッジルーターへ至る複数のラベルスイッチパスを設定することにより、当該負荷分散装置を始点とするこれら複数のラベルスイッチパスの間で、負荷分散を行うことが可能である。これにより、任意のラベルスイッチパスの品質が経路途中で劣化した場合、自動的に他のラベルスイッチパスへフローを振り向けることが可能となり、常にネットワークの利用効率を高く保つことができる。

【００６０】また、図１０では、ＰＰＰの仮想コネクションで経路を区別する場合を示している。ＰＰＰはＯＳＩ参照モデル第２層のプロトコルであり、その仮想コネクションは一般に２つの隣接するノード間を接続するものである。任意のノードが複数のＰＰＰコネクションの終点になっている場合、一般には各仮想コネクションに属するパケットを明確に識別することが可能である。すなわち、各仮想コネクションごとに伝送媒体が異なる場合、あるいはＰＰＰｏＥ（PPP over Ethernet（登録商標））のように、同一の伝送媒体を介して複数の仮想コネクションが設定されている場合には、仮想コネクションごとに固有の｛送元ＭＡＣアドレス、宛先ＭＡＣアドレス、セッション識別子｝の組が割り当てられる場合もある。

【００６１】このようなＰＰＰ仮想コネクションを使用した負荷分散の例として、１つの顧客が同時に複数のサービスプロバイダに接続し、これらを経由してネットワークへ接続する場合を例として説明する。以下では、説明を簡単にするため、顧客と各サービスプロバイダとの間をそれぞれ１つのＰＰＰ仮想コネクションで接続する場合を想定する。すなわち、顧客が所有する上記負荷分散装置は、それぞれ異なるＰＰＰ仮想コネクションを介して複数のサービスプロバイダのリモートアクセスサーバーとそれぞれ接続しているものとする。

【００６２】上記ケースでは、各サービスプロバイダに対して１つのグループを割り当て、それぞれのプロバイダへ向かう使用帯域幅に応じて、対応する重み係数の比を調節することができる。すなわち、一方のサービスプロバイダが混雑しており、他方のサービスプロバイダが空いている場合には、混雑しているサービスプロバイダに向かうフローを、空いているサービスプロバイダの方へ自動的に振り向けることができる。これにより、複数のサービスプロバイダの混雑状況が時々刻々変化する状況においても、各サービスプロバイダに対して効率よくパケットを転送することが可能となる。

【００６３】図１１では、ＡＴＭのＶＰまたはＶＣを使用する場合を示している。ＡＴＭのＶＰまたはＶＣは、例えばルーター間を接続するために用いることができる。以下では、上記負荷分散装置から任意の同一宛先へ向かう２つの経路が存在し、各経路の次のホップのルーターがそれぞれ異なるＶＰまたはＶＣにより接続されている構成を例として説明する。上記ケースでは、負荷分散装置において当該同一宛先へ向かうトラフィックを２つのグループに分類し、それぞれの使用帯域幅を測定した結果に基づき、各グループに対する重み係数を調整することができる。

【００６４】これにより、例えば一方のＶＰまたはＶＣに障害が発生して、同一宛先に到達できなくなっても、自動的に他方のＶＰまたはＶＣに対する重み係数が増加

するため、障害のある一方の経路から障害のない他方の経路へフローを移動させることができる。なお、上記同一宛先は、必ずしもパケットの終着点に限られるわけではなく、例えば途中経由するネットワークへの経路が複数存在する地点を同一宛先と見なすこともできる。

【００６５】図１２では、ＩＥＥＥ８０２．１ＱのＶＬＡＮを用いて経路を区別する場合を示している。例えば、複数のＶＬＡＮを経由して同一宛先へパケットが到達可能である場合、各ＶＬＡＮをそれぞれ１つの経路として扱うことができる。この場合、各経路に対して重み係数を割り当てて、各ＶＬＡＮを経由するトラフィックを測定し、この結果に基づき各重み係数を調節することにより、各ＶＬＡＮを効率的に利用してトラフィックを転送することが可能となる。

【００６６】図１３では、次のホップのノードに基づき経路を区別する場合（物理的にインターフェイスが異なる場合）を示している。ここでいうノードとは、ネットワーク層（あるいはデータリンク層）の情報をもとにパケットを送受信あるいは転送する機能を持つもの一般を指す。次のホップとは、送元から宛先へパケットを転送する経路上の任意のノードから見て、次に経由するノードのことである。以下では、インターネットにおいて、任意のサービスプロバイダのルータがＩＸ（Internet eXchange）を介して、他の複数のサービスプロバイダに接続している場合を例として説明する。

【００６７】パケットが同一宛先ネットワークに到達するための次ホップとして、複数のサービスプロバイダーを選択可能である場合には、これら複数のサービスプロバイダを経由する経路に対して、本発明にかかる経路制御方法を用いて負荷分散を行うことが可能である。ＩＸの構造が、データリンク層のコアスイッチを取り囲んで各プロバイダのルーターが接続されている構造をなしている場合、経路を識別する具体的な手段としては、次ホップのノードのＩＰアドレスと、これに対応するデータリンク層アドレスを用いることができる。

【００６８】上記宛先ネットワークへ宛てたパケットを、これら複数経路へのグループに分類し、それぞれのグループに対する使用帯域幅を測定して、重み係数の値に反映させることにより、各経路の品質が時々刻々変化する場合においても、常に各経路の資源を有効に利用して、トラフィックを転送することが可能となる。例えば、任意のサービスプロバイダを経由する経路の品質が劣化した場合には、そこを通るフローを直ちに他のサービスプロバイダを経由する経路へ自動的に振り向けることができる。逆に、任意のサービスプロバイダを経由する経路が他の経路よりも空いた場合には、他の経路から空いている当該経路へ自動的にフローを振り向けることができる。

【００６９】また、任意の経路の途中のプロバイダに障害が発生して宛先ネットワークへ到達不能となった場合

は、その経路を経由するＴＣＰのコネクションにおいて輻輳制御機能が働き、急速にスループットが劣化するため、これを測定して重み係数を更新することにより、その障害を速やかに回避することが可能となる。ここで、特に重要なのは、このような負荷分散機能や障害回避機能を、上記負荷分散装置が自律的に行うことである。すなわち、他のルーターやサーバーなどから、経路情報や障害情報をいっさい取得せずに、負荷分散装置を通過するトラフィックから得られる情報のみで経路制御を行うことができる。

【００７０】上記した例では、単一プロバイダが上記負荷分散装置を使用する場合について説明したが、複数のプロバイダが並列的にそれぞれの負荷分散装置を用いてもよい。さらに、上記例では、１つの宛先についての負荷分散処理を例として説明したが、任意の宛先への負荷分散処理を行いながら、同時に別の宛先への負荷分散処理を行うことも可能である。

【００７１】また、上記負荷分散装置がトラフィックを分散する対象は、同種の経路に限定されるものではない。例えば、同一宛先へ到達することができる３つの経路が存在し、その第１の経路はＭＰＬＳのラベルスイッチパスを使用し、第２の経路はＰＰＰの仮想コネクションを使用し、第３の経路はＡＴＭのＶＣを使用して他のルーターに接続している場合においても、これら経路間で負荷分散を行うことも可能である。その理由は、本発明にかかる経路制御方法および装置が、任意の複数経路に対して、上記負荷分散装置を通過するトラフィックの使用帯域幅を測定できれば、各経路の具体的手段には依存しないからである。

【００７２】次に、図１４を参照して、フロー分類部１でのフローの管理方法について説明する。図１４はフロー管理方法を示すフローチャートである。本実施の形態では、重み係数の比Ｗ１１：Ｗ１２が変化した時に、経路の間でのフローの移動が発生する。このとき、２つの経路の間で遅延時間に差がある場合、同一フローに属するパケットの順序が、途中で入れ替わる可能性がある。例えばＴＣＰなどを利用している場合には、同一フロー内でのパケット順序の入れ替えはパフォーマンスの劣化の原因となる可能性がある。

【００７３】そこで、図１４に示すように、フロー分類部１において、重み係数が変化した場合でも同一フローに属するパケットが必ず同一経路を通るように管理する。この方法では、フロー記憶テーブル７を設けて、通過するパケットが属するフローとそのフローに割り当てられている経路を対応付けて登録する。まず、パケットが入力すると、フロー記憶テーブル７を参照してパケットが属するフローが既に登録されているかどうかを検索する（ステップ１１０）。その結果、フローが既に登録されている場合には（ステップ１１１：ＹＥＳ）、そのフローに対応するグループにパケットを分類する（ステ

ップ１１２）。一方、フローが登録されていない場合には（ステップ１１１：ＮＯ）、重み係数を増加させたいグループにパケットを分類し、対応するフローとグループとを新たにフロー記憶テーブル７に登録する（ステップ１１３）。

【００７４】図１５に、フローとしてＴＣＰコネクションを使用した場合のフロー記憶テーブル７の例を示す。フローは送元アドレス、宛先アドレス、送元ポート番号、宛先ポート番号の４つの数を識別子として識別し、これに対して出力クループが対応付けられている。この例では、フロー識別子が｛ＳＩＰ１，ＤＩＰ１，ＳＰＯＲＴ１，ＤＰＯＲＴ１｝なるパケットはＧ１１に分類され、フロー識別子が｛ＳＩＰ２，ＤＩＰ２，ＳＰＯＲＴ２，ＤＰＯＲＴ２｝なるパケットはＧ１２に分類される。入力パケットのフロー識別子が、これらのいずれかに一致する場合には、対応する出力グループに分類される。また、入力パケットのフロー識別子が、これらの何れにも一致しない場合には、重み係数を増加させるべきグループにパケットを分類することができる。

【００７５】図１６に登録されていないフローをどのグループに登録するかを判断する際に用いることができるフロー割り当てテーブル８の例を示す。このフロー割り当てテーブル８は、各グループに対して現在の重み（登録されているフロー数の現在値）と、目標とする重み（登録フロー数の目標値）が対応付けられている。目標とする重みは、そのときの重み係数の比に等しくなるように計算することができる。例えば、重み係数の比Ｗ１１：Ｗ１２が１：２であると仮定し、フロー識別子｛ＳＩＰ１，ＤＩＰ１，ＳＰＯＲＴ３，ＤＰＯＲＴ１｝なるパケットが入力した場合には、対応するフロー識別子は図１５のフロー記憶テーブル７に登録されていない。そこでフロー割り当てテーブル８を参照すると、Ｇ１２の現在の重みが目標とする重みよりも小さいことが分かる。

【００７６】したがって、このパケットをＧ１２に登録するとともに、新しいフロー識別子｛ＳＩＰ１，ＤＩＰ１，ＳＰＯＲＴ３，ＤＰＯＲＴ１｝を、グループＧ１２と対応付けてフロー記憶テーブル７に登録し、さらにフロー割り当てテーブル８の、グループＧ１２に対応する登録フロー数を１増加する。例えば、現在のインターネットでは、フローの平均持続時間は数秒であり、上述の方法を逐次行うことにより、同一フローに属するパケットの順序入れ替えが生じることなく、重み係数に比例したフロー数を各経路に出力することが可能となる。なお、フロー記憶テーブル７は、エージングタイマーを用いて、古くなったフローの登録を削除するようしてもよい。

【００７７】

【発明の効果】以上説明したように、本発明は、入力されたパケットのフローを予め設定されている各グループの重み係数に基づきいずれかのグループへ分類し、各グループのフローを当該グループごとに異なる経路へ出力するものとし、各経路で使用されている帯域幅のうち各フローで使用されている帯域幅をグループごとに測定して、これら帯域幅の比に基づき各グループの重み係数を調節するようにしたものである。したがって、各経路を通るフローで使用される帯域幅が平均化されることから、ネットワーク内の複数経路にわたるトラフィックの配分を動的に最適化でき、ネットワーク資源を効率的に利用することができる。

【図面の簡単な説明】

【図１】　本発明の一実施の形態にかかる経路制御方法が適用されるパケット転送システムを示す概略構成図である。

【図２】　複数経路を示す概念図である。

【図３】　負荷分散装置の構成例を示すブロック図である。

【図４】　負荷分散装置での経路制御処理を示すフローチャートである。

【図５】　重み係数の調整方法例を示す説明図である。

【図６】　ＯＳＩ参照モデルの第４層コネクションをフローとみなす場合のパケット分類例である。

【図７】　ＯＳＩ参照モデルの第３層アドレスをフローとみなす場合のパケット分類例である。

【図８】　ハッシュ関数でパケットを分類するフロー分類部の構成例である。

【図９】　ＭＰＬＳネットワークのＬＳＰで経路を区別する場合のパケット出力例である。

【図１０】　ＰＰＰの仮想コネクションで経路を区別する場合のパケット出力の例である。

【図１１】　ＡＴＭネットワークの仮想コネクションで経路を区別する場合のパケット出力の例である。

【図１２】　ＶＬＡＮネットワークのＶＬＡＮ識別子で経路を区別する場合のパケット出力の例である。

【図１３】　出力物理インターフェイスで経路を区別する場合のパケット出力の例である。

【図１４】　フロー管理方法を示すフローチャートである。

【図１５】　フロー記憶テーブルの構成例である。

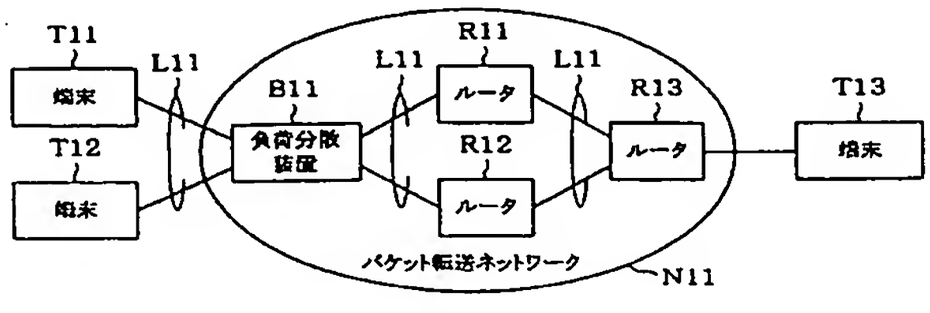
【図１６】　フロー割り当てテーブルの構成例である。

【符号の説明】

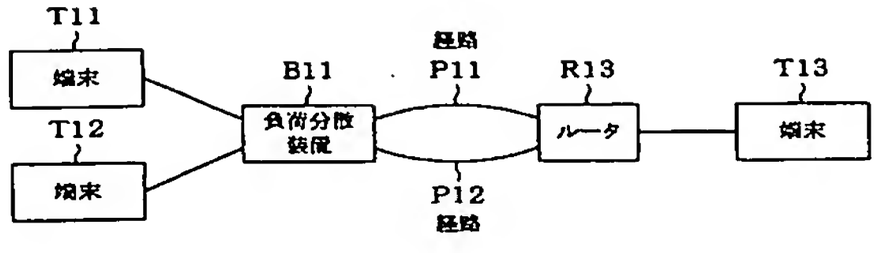
Ｎ１１…パケット転送ネットワーク、Ｂ１１…負荷分散装置（経路制御装置）、Ｒ１１，Ｒ１２，Ｒ１３…ルータ、Ｔ１１，Ｔ１２，Ｔ１３…端末、Ｐ１１，Ｐ１２…経路、Ｇ１１，Ｇ１２…グループ、Ｗ１１，Ｗ１２…重み係数、ＢＷ１１，ＢＷ１２…単位使用帯域幅、Ｎ１２…ＭＰＬＳネットワーク、Ｎ１３…ＰＰＰネットワーク、Ｎ１４…ＡＴＭネットワーク、Ｎ１５…ＶＬＡＮネットワーク、Ｎ１６…物理的に経路が分かれたネットワーク、１…フロー分類部、２，３…フロー出力部、４…

帯域幅測定部、５…単位帯域幅算出部、６…重み係数調整部、７…フロー記憶テーブル、８…フロー割り当てテーブル。

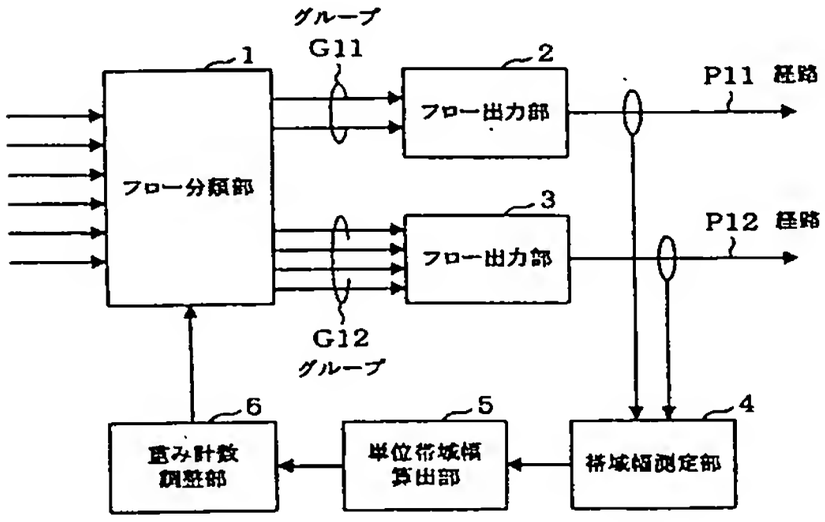
【図１】



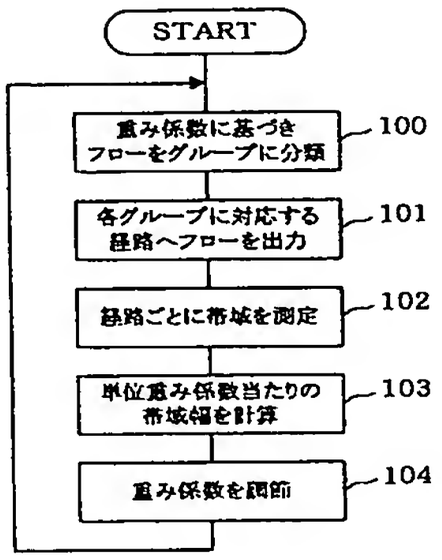
【図２】



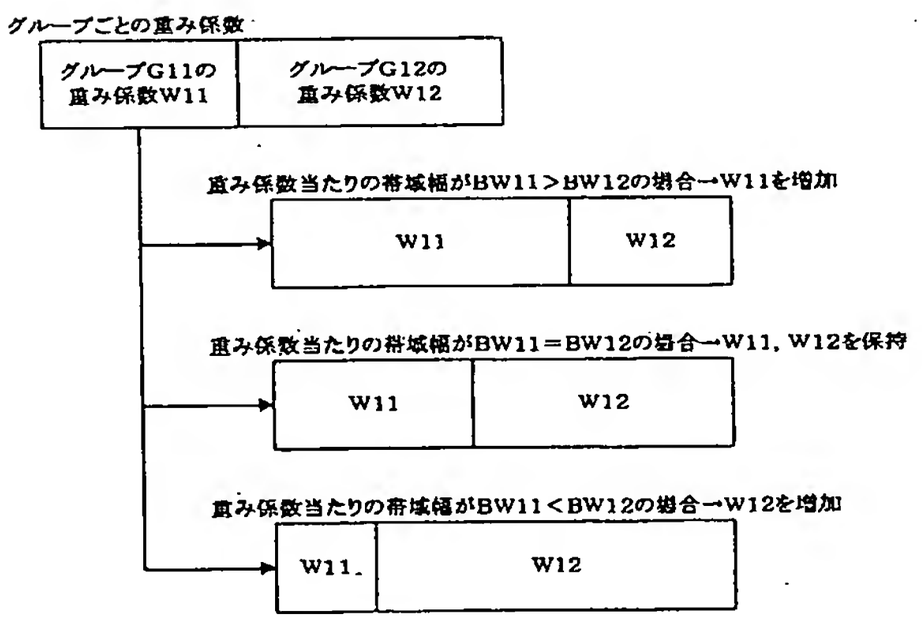
【図３】



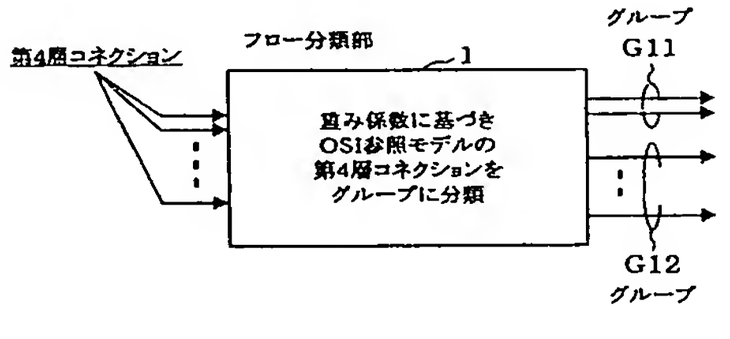
【図４】



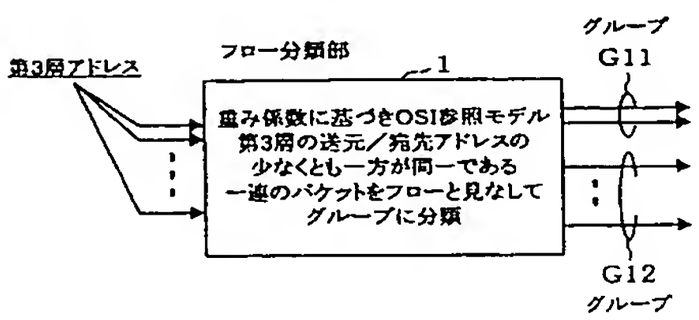
【図５】



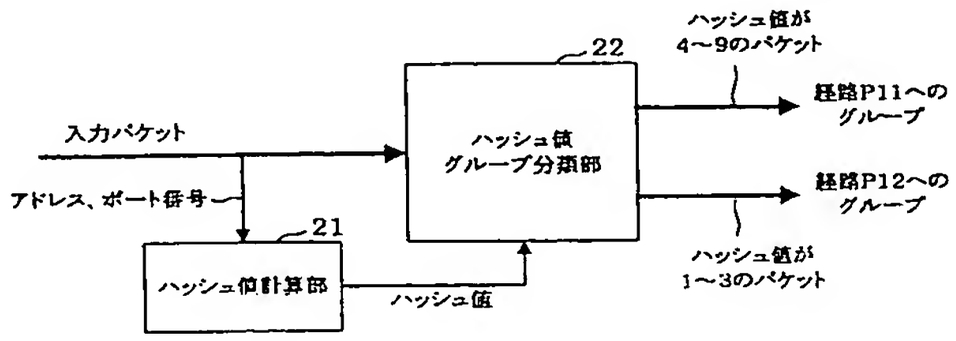
【図６】



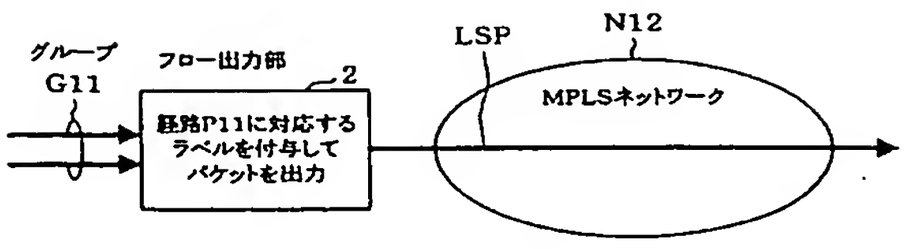
【図７】



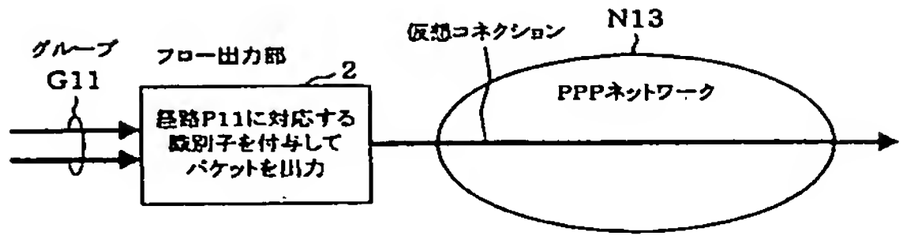
【図８】



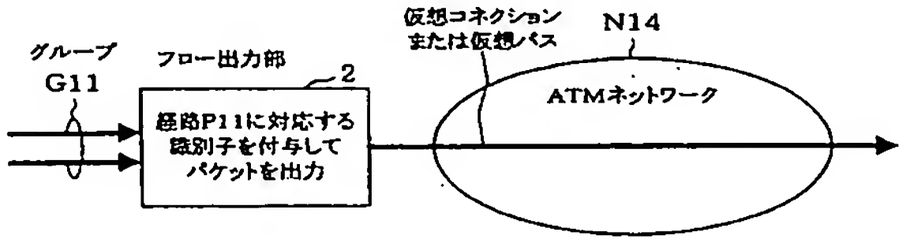
【図９】



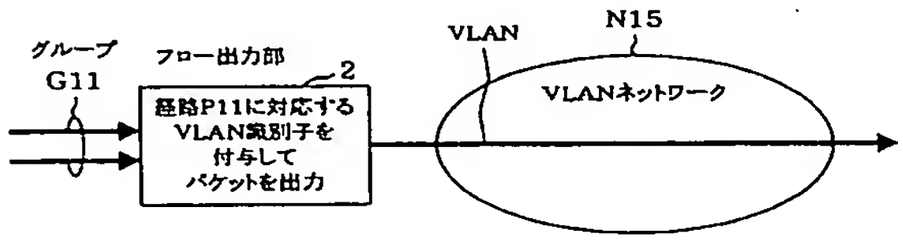
【図１０】



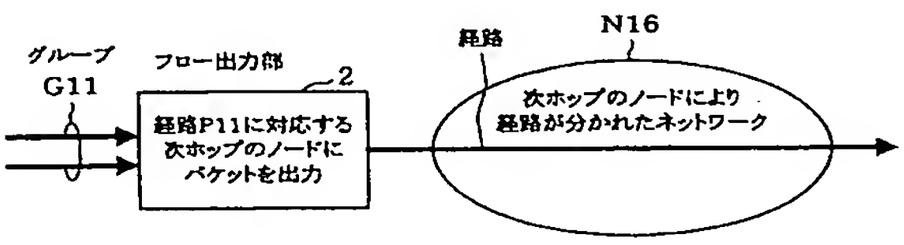
【図１１】



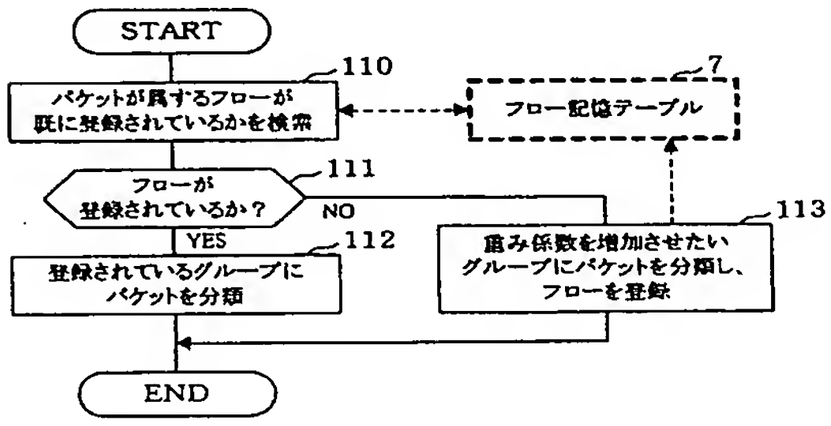
【図１２】



【図１３】



【図１４】



【図１５】

フロー記憶テーブル　7

フロー識別子	グループ
送元アドレス　=SIP1 宛先アドレス　=DIP1 送元ポート番号　=SPORT1 宛先ポート番号　=DPORT1	G11
送元アドレス　=SIP2 宛先アドレス　=DIP2 送元ポート番号　=SPORT2 宛先ポート番号　=DPORT2	G12

【図１６】

フロー割り当てテーブル　8

グループ	現在の重み （登録フロー数の現在値）	目標とする重み （登録フロー数の目標値）
G11	1	1
G12	1	2